The Pennsylvania State University

The Graduate School

College of the Liberal Arts

INFANT VISUAL HABITUATION MODELED BY NON-LINEAR

REGRESSION

A Thesis in

Psychology

by

Michael P. Dahlin

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

May 2004

The thesis of Michael P. Dahlin was reviewed and approved* by the following:

Hoben Thomas
Professor of Psychology
Thesis Advisor
Chair of Committee

Rick O. Gilmore
Assistant Professor of Psychology

Keith Nelson
Professor of Psychology

Tom Hettmansperger
Professor of Statistics

Kevin Murphy
Professor of Psychology
Head of the Department of Psychology

*Signatures are on file in the Graduate School

ABSTRACT

For nearly 40 years, visual habituation techniques have provided insight into infant perceptual and cognitive development, although relatively little attention has been given to the development of formal mathematical models of the habituation process. Yet model-based measures outperform conventional habituation criteria (the fifty percent decrement criterion, or 50%DC) in several important ways. Model-based approaches better distinguish between infants whose looking time responses exhibit systematic decreases over trials and infants who do not. Use of model-based measures provide increased sensitivity for the detection of post-habituation recovery responses to novel stimuli, even when conventional methods based upon the 50%DC fail. Finally, model-based measures can be implemented using fixed numbers of habituation trials (as few as seven), eliminating the need for online computation of habituation criteria between trials.

The current study used a conventional infant control habituation paradigm with 55 infants (aged 4-6 months) to observe pre- and post-habituation responses to one of three pairs of visual stimuli. Infants were repeatedly presented with one stimulus from the pair for up to a maximum of 15 trials, or until they achieved the 50%DC. Infants then saw six additional trials, alternating between the "habituated" stimulus and a similar but perceptually

distinguishable novel stimulus. Roughly one week later, infants were tested a second time using the same stimulus pair, but reversing their roles such that the habituating stimulus from the first visit became the novel stimulus on the second visit. Looking time responses for each infant visit were classified under one of four habituation models using BIC (Ramsey & Schafer, 1997, p. 343-345), and the magnitudes of post-habituation recovery were examined using both conventional and model-based measures. Infant habituation behavior (i.e., model classification) was found to be independent across visits. Furthermore, habituation behavior and post-habituation recovery were independent of stimulus condition, even though the stimulus pairs did not elicit equivalent interest levels among infants.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

Chapter 1

Visual Habituation as a Behavioral Tool

Visual habituation in infants refers to their tendency to show decreasing attention towards stimuli presented repeatedly over time. Operationally, one might define visual habituation as the tendency for infants' looking times towards an object or stimulus to decline from an initially higher level to an asymptotic floor value as the number of presentations, or trials, grows increasingly large. Figure 1 shows this relationship for a four-month old infant from the current
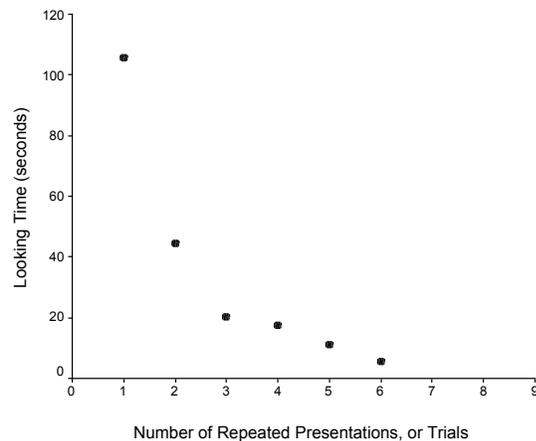
Figure 1: Looking Times Over Repeated Presentations of a Stimulus

study, shown a continually repeating film loop of a smiling baby. This infant looked at the moving image for nearly two minutes the first time it was shown to her, but declined gradually to about six seconds by the sixth trial.

While this kind of attenuation of interest is not exclusive to infants, it has provided a popular tool for examining questions regarding perceptual and cognitive awareness during infancy (Bornstein, 1989). Measuring and observing a decline in an infant's looking time to repeated images of a kitten, for example, then finding a subsequent renewal of interest towards the image of a horse implies that the infant perceptually distinguishes between the two types of images. This example, albeit simplistic, provides a general sense of a method that can be applied to a multitude of sensory stimuli and/or perceptual categories, and is sometimes extended even further to make inferences about infants' understanding of the underlying concepts represented by the visual stimuli (e.g., Arterberry & Yonas, 2000; Baillargeon, 1991; Bornstein, 1989; Johnson & Aslin, 2000; Kagan, Linn, Mount, Reznick & Hiatt, 1979; Kirkham, Slemmer & Johnson, 2002).

Brief History

Although habituation techniques have grown increasingly popular since the 1960s as tools for better understanding infant perceptual and cognitive

development, they were initially used to study learning by researchers who observed preferences for novel (vs. familiar) stimuli in species as diverse as human beings, rats and paramecia (Cohen, 1976). Neural mechanisms have been proposed (e.g., Sokolov, 1963; Sokolov, Spinks, Näätänen & Lyytinen, 2002) for such novelty preferences among higher organisms, hypothesizing relationships between the strength of "internal mental representations" of the presented visual stimuli and their corresponding orienting responses (i.e., looking times) towards the stimuli. Under such theories, decreased looking times to repeated stimuli imply a strengthening of the mental representation, or increasing correspondence between the external stimuli and the internal representations. Additional theories posited concurrent relationships between the underlying processes driving habituation and specific cognitive abilities (e.g., Feldman & Mayes, 1999; Millar & Weir, 1997), as well as predictive relationships between habituation behavior and subsequent cognitive performance (e.g., Bornstein & Tamis-Lemonda, 1994; Fagan, 1984; Miller, Ryan, Short, Ries, McGuire & Culler, 1977).

Fantz's (1964) examination of novelty preferences is widely held to mark the beginning of the modern era of infant habituation research (Bornstein, 1985a). Fantz developed a "fixed trial" technique in which he measured infants' preferences for novel vs. familiar stimuli by examining corneal reflections to determine their directions of gaze (Cohen, 1976). He found no consistent novelty preference in infants younger than two months, but did find systematic novelty

preferences after about four months of age. Several studies specifically examining visual habituation followed suit, all using Fantz's fixed trials procedure (Cohen, 1976). These studies showed habituation behavior in infants as young as ten weeks old.

Habituation Paradigms

Scores of studies using visual habituation techniques began to appear in the literature thereafter (over 200 in the past forty years, according to a recent PsychInfo search using the terms "infant visual habituation"; hundreds of others appear when using more general search parameters), and three general classes of habituation methods emerged to supplement Fantz's original fixed trial procedure: fixed level, free-looking, and infant control procedures (Bornstein, 1985a). Under the fixed trial procedure, each infant is presented with a fixed number of habituation trials, each trial of a constant duration, with a short break or inter-stimulus interval possibly occurring between presentations. Specific examples include Fantz's (1964) original study using ten trials of sixty second durations and a brief but unspecified inter-stimulus interval, Dodd and Lewis' (1969) study using seven thirty-second trials separated by thirty-second inter-stimulus intervals, and Caron and Caron's (1969) study using fifteen trials of twenty-second durations with no inter-stimulus interval. Under this paradigm, potential dependent measures included the proportion of time per trial that the infant gazed upon the stimulus, or the total looking time across all trials. Under

the second class of habituation procedures, fixed *level* procedures, trial durations were still held constant, but the number of trials varied across infants. Under this variation, infants viewed repeated trials of the stimulus until their gazes met some fixed absolute habituation criterion such as looking at the stimulus for no more than three seconds on two consecutive trials, as in McCall, Hogarty, Hamilton and Vincent (1979). The dependent measure of focus under this paradigm was typically the number of trials to criterion (Bornstein, 1985). Under the third category, called free-looking procedures, the familiar stimulus is shown to the infant for a single time period. During this single, usually longer interval (e.g., 180 seconds, as in Bornstein, Ferndinandsen & Gross, 1981), researchers might measure the number of individual instances, called fixations, that the infant gazed at the stimulus, or the proportion of time during which the infant looked at the stimulus. Under the final class of habituation procedures, called infant control procedures, the durations of individual habituation trials vary, subject to the infant's "control". Under this variation, trials continue for as long as infants gaze at the stimulus, ending when they look away. Originally, infant control procedures utilized a fixed number of habituation trials before presenting the novel stimulus, (Horowitz, Paden, Bhana & Self, 1972b), but the term soon began to refer to methods using variable numbers of trials subject to a relative behavioral criterion (e.g., Boyd, 1974; Laub & Bhana, 1974).

These researchers noted that infants exhibited wide ranges in their minimal attention levels, and that use of absolute criteria were problematic for individuals whose minimal looking times tended to be much higher or lower than the arbitrary fixed value, causing fatigue or frustration which reduced their likelihood of completing the experiment. Furthermore, some infants remained docile and willing to participate for many more trials than others, making it difficult to determine the most appropriate fixed number of trials to use. Alternatively, researchers computed a relative criterion using the observed looking times from the first two (e.g., Bornstein & Tamis-LeMonda, 1994), three (e.g., Baillargeon, 1991) or four (e.g., Kirkham, Slemmer & Johnson, 2002) trials as a baseline against which to compare subsequent observations. All these studies used a variation of what Dannemiller (1984) called the fifty percent decrement criterion (or 50%DC), in which a rolling average computed from the looking times for the last few trials is compared to an initial baseline value generated from the first few trials. Once this rolling average is less than or equal to fifty percent of the baseline value, habituation trials are terminated. Using the notation $y_i(t)$ to refer to an infant i's observed looking time response for any observed trial t (please see Appendix A for a complete glossary of notation and symbols used within the manuscript), one would compute the baseline average looking time, or $\bar{y}_{bi}$ (using, for example, three trials) as shown in Equation 1:

$$\overline{y}_{bi} = \frac{\sum_{t=1}^{3} y_i(t)}{3} \tag{1}$$

In such studies, habituation trials continue until the infant's current rolling

average $\overline{y}_{ci} = \frac{y_i(t_c) + y_i(t_{c-1}) + y_i(t_{c-2})}{3} \leq \frac{\overline{y}_{bi}}{2}$, where $y_i(t_c)$ refers to the observed looking

time for the current trial.

Of the four major categories or types of habituation procedures, fixed trial

and infant control procedures have been the most popular, although criticisms

have been leveled against both (Bornstein, 1985a). Fixed trial procedures,

because they control both the number and duration of habituation trials, assume

that all infants habituate at relatively the same rate and to the same degree,

limiting their utility as measures of individual differences in habituation response

behavior (Lewis, 1969). Infant control procedures were designed to better

accommodate such variability among infants, providing measures of individual

differences and reducing participant attrition (Horowitz, Paden, Bhana & Self,

1972b; Laub & Bhana, 1974). Such procedures, however, are more procedurally

complex as they require the calculation of the relative criterion between each

habituation trial to determine the termination of the habituation phase within an

experimental visit (Bornstein, 1985a).

Even with their greater methodological complexity, infant control

procedures using variable numbers of habituation trials and the 50% decrement

criterion are the most common procedures currently in use (Bornstein, 1998).

The 50%DC has gained widespread popularity for its ease of computation and for

its surface validity: It can be computed for all infants, regardless of individual

variation in their maximal and minimal looking time responses for any given

stimulus.

Shortcomings exist with this criterion, however. In particular, the 50%DC

criterion is problematic for infants whose minimal looking time responses are

elevated above zero. In their Footnote 5, Thomas and Gilmore (in press)

demonstrate the increasing tendency for the 50%DC to be non-achievable as an

infant's asymptotic minimal, or floor response level increases. As a case in point,

consider the following hypothetical situation. Although an infant's "true" (i.e.,

error free) habituation curve cannot be precisely known, one can imagine an

infant whose observed responses resemble those in Figure 2, with a true initial
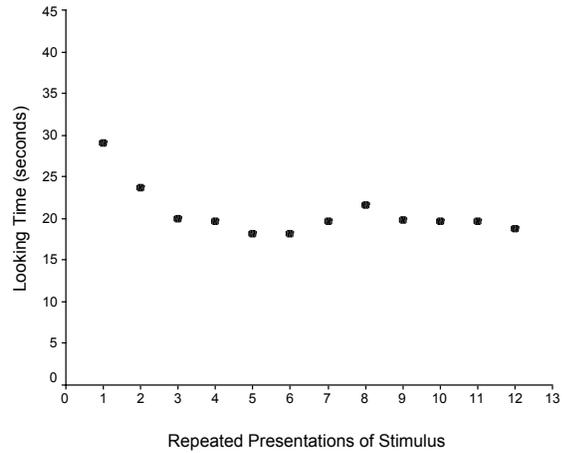
Figure 2: Infant With a High Floor Value: Max. $\approx$ 30 Sec., Min. $\approx$ 20 sec.

level at roughly 30 seconds and a true minimal attention level at about 20 seconds in response to repeated presentations of a visual stimulus. Such a child would be less likely to achieve the 50%DC because that child's observed baseline response level, $\bar{y}_{bi}$, based on the first three trials, is approximately 24. A baseline of 24 corresponds to a 50%DC of 12, which is lower than that infant's potential range of responses. Similarly, one can imagine another child with an initial attention level of 45 seconds and an asymptotic minimal attention level of 2 seconds, as depicted in Figure 3. For this infant, $\bar{y}_{bi} \approx 22$ and the 50%DC ≈ 11.
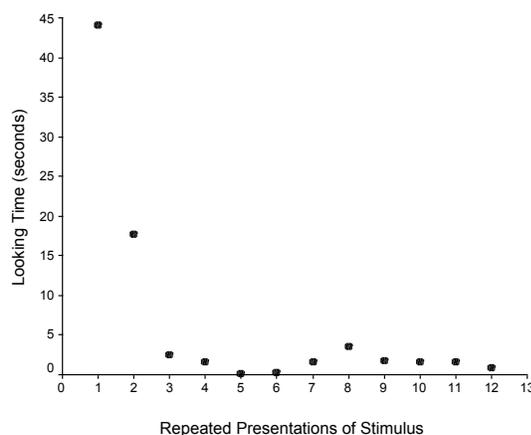
Figure 3: Infant With a Low Floor Value: Max. $\approx$ 45 Sec., Min. $\approx$ 2 Sec.

In the first example, then, the 50%DC represents a point well below that

individual's floor while in the second example, the 50%DC represents a point

somewhere above that infant's asymptotic floor. If the intended goal of the

50%DC is to identify the trial at which infants have achieved proportionally

equivalent states of habituation, relative to maximal and minimal response levels

that vary across infants, such as Colombo implied (1993, p. 39), then these two

examples show how the criterion fails. Unless one assumes that all infants share

a common, low floor value (i.e., at zero seconds) relative to their initial levels,

infants will not be brought to "equalized" states of habituation because each

infant's looking time responses will not have decreased by the same proportion.

Another shortcoming with this criterion is the fact that it relies on observed looking time measurements without considering the potential influence of random error on observations. Without specifying some formal parametric model and then estimating the parameters of that model, there is no way to assess how much any given observation differs from what might be expected on that particular trial, or how such error within observations may be influencing various outcomes (e.g., attaining the 50%DC). Indeed, the notion that individuals might achieve habituation criteria without actually habituating (i.e., because of random error rather than systematic habituation) cannot be adequately addressed or meaningfully discussed without imposing some formal structure or conceptualization of the habituation process and its corresponding error.

Several researchers (e.g., Cohen & Menten, 1981; Dannemiller, 1984; Ashmead and Davis, 1996) have sought to overcome this shortcoming, proposing mathematical models of the habituation process in order to examine the effectiveness of commonly used habituation measures, and to examine how error variance in observed responses may affect the validity of such measures.

Modeling of Habituation

Although infant-control procedures were an improvement over fixed trial procedures in their reflection of individual differences, some researchers have questioned the validity of the stopping criteria commonly used under this paradigm to determine when habituation has been achieved (e.g., Dannemiller, 1984; Ashmead & Davis, 1996). Two issues of particular concern focus on the impact of error on observed responses, and the notion that infants are actually brought to equalized states of habituation under this criterion. A model-based approach provides analytical tools for examining both issues. With respect to the first issue, models can be used to explore the relative impact of error on the validity of the stopping criterion under varying initial conditions through computer simulation. With respect to the second issue, that of equalizing infants, real data can be evaluated against both formal habituation models and conventional habituation criteria such as the 50%DC to determine which, if either approach produces more consistent "post-habituation" responses across groups of infants.

Simulation studies under specific models have also been used to understand group-based habituation data, as well as to highlight the important differences between individual behavior and group-based habituation curves. Cohen and Menten (1981), for example, compared two simulations under an

exponentially decreasing model and an "all-or-none" model, generating group

mean looking times for any particular trial t under both models in the absence of

any error term. In their first equation, represented here by Equation 2, they

posited that a group mean looking time on a trial j, μ_j, was computed by the

following exponentially decreasing model:

$$\mu_j = \exp(-jr)(\mu_o - \mu_f) + \mu_f \tag{2}$$

where μ_o is the group mean looking time for trial 1 ($\mu_o \geq 0$), μ_f is the group mean

looking time on the final trial ($\mu_f \geq 0$), and r is a positive slope parameter

determining how quickly the mean looking time decreased over trials. This

model specified monotonic decline in the group mean over all increasing trials j,

such that $\mu_j \geq \mu_{j+1}$.

Their all-or-none model, represented here in Equation 3 expressed the

group mean at trial j, μ_j, as a weighted combination of the proportion of infants

responding at the initial response level μ_o, and the remaining group whose

responses had dropped to a final level, μ_f. Under this model, infants' mean

responses are at the initial level μ_o on trial one and drop to μ_f on any trial j with

probability p (where $0 \leq p \leq 1$). This, too, is a monotonically decreasing model

such that once an infant's response level has dropped, it remains at μ_f for all

subsequent trials.

$$\mu_j = (1-p)^j(\mu_o - \mu_f) + \mu_f \tag{3}$$

Simulated data produced by both these models were compared to group mean data from an empirical sample. They claimed that the simulations under these models generated mean curves that were indistinguishable from each other and similar to observed sample data, although no objective goodness of fit or other statistical measures were provided. Furthermore, they claimed that specific characteristics of group-based habituation curves may in fact be artifactual, concluding that individual differences measures, rather than group-based models, would be preferable for the task of drawing inferences about specific internal perceptual processes within infants.

Given their decision not to provide goodness of fit or other objective statistical measures comparing the two models, one cannot know how well they might work with real data. Nevertheless, the study was an important contribution in that it highlighted the potential problems associated with using group-based measures to draw inferences about individual behaviors.

Dannemiller (1984) used a model-based approach to explore measurement issues associated with individual rather than group-based habituation curves. Specifically, he looked at how random error affected the validity of the 50%DC in a simulation using an exponentially decreasing model with fixed parameter values and varying error magnitudes. Dannemiller's approach began by specifying an exponential model for $X_a(t)$ with three fixed parameters: $X_a(t) = (g)\exp(\text{-}rt)+k,$

where X_a was a non-negative continuous random variable denoting the "actual"

value at trial t. g (≥ 0) was a fixed value denoting the model's range, r (≥ 0)

denoted the model's slope and k (≥ 0) was the final floor value. To this fixed

model a mean zero random normal error term $E(t)$ was added, such that $X_o(t) =$

$X_a(t) + E(t)$, where $X_o(t)$ denoted the "observed" value. Dannemiller then

rescaled the observed values under the model by dividing each $X_o(t)$ within a

simulation by the known $X_a(t{=}1)$. Though his motivation for doing so was not

specified, this apparently permitted comparisons of simulations with differing

fixed values of g, since the rescaled model values would then become independent

of g. This rescaled model, Dannemiller's Equation 6, is reproduced here as

Equation 4:

$$X_o(t) = (1 + e)[(1 - k)\exp(-r(t - 1)) + k] \tag{4}$$

where X_o denotes the observed relative looking time response expressed as a

proportion (where $0 \leq X_o \leq 1$) and t denotes the trial number. e denotes a

relative error term in which $e(t) = E(t)/X_a(t)$. The rescaled k denotes the

minimal response time expressed as a proportion of the maximum response (such

that $0 \leq k \leq 1$), and r denotes a slope, or rate of habituation.

Using fixed parameter values and varying error magnitudes, his

simulations found that even marginal levels of error strongly impacted the

number of trials necessary to achieve the 50% decrement criterion. Additionally,

and perhaps more importantly, he found that with sufficient magnitudes of error, the 50%DC was achievable even for non-habituating individuals (i.e., simulations in which r was specified to be nearly zero under his model), given as few as six trials.

Dannemiller's scaled simulation model expressed outcomes as proportions to allow comparisons between simulated infants with differing maximal and minimal response times. Thus independent parameters for the maximum and minimum response times were not part of the scaled model. Such a model would be problematic for use with real data, however, since the true scores and error magnitudes upon which each observation would be scaled are not known in advance. Parameter estimates under this model would therefore not be directly identifiable.

Even though Dannemiller's model was not evaluated against real data, his work highlighted an important consideration for those who would use an infant control paradigm in their research: the notion that an infant might demonstrate a 50% reduction in looking time from an initial baseline level because of random variation rather than any systematic habituation. While surely no one would disagree that observed habituation data contains random error attributable to various sources, the notion that such errors may lead to what he termed "false alarms", or erroneous achievement of the habituation criterion, seems to have

been largely ignored by proponents who believe that under infant-controlled procedures, "...every infant tested actually habituates..." (Colombo, 1993, p. 42).

Dannemiller's findings were consistent, too, with Cohen and Menten's (1981) conclusions regarding the differences between individual and group-based response measures. Specifically, Dannemiller showed that a given criterion (e.g., the 50%DC) might be achieved in two ways: by actual habituation and by random fluctuation. Thus, conventional data analytical approaches such as the paired t-test are ill equipped for situations in which multiple groups or strategies give rise to observed differences (Thomas & Gilmore, in press). A consequence is that the responses of individuals within observed groups may not be described by a single underlying probability distribution, hence those responses are not identically distributed, which is a critical assumption of conventional group-based analyses. Consequently, such analyses may provide misleading conclusions.

Though Dannemiller (1984) identified shortcomings with the 50%DC, he proposed no alternatives. Ashmead and Davis (1996) did precisely this with another approach, comparing the conventional 50%DC (using baselines computed from 2, 3 or 4 trials) with an alternative criterion based on a best-fit polynomial regression equation computed individually for each simulated infant i of the form shown in Equation 5:

$$y_i(t) = a_i + b_{i1}t + b_{i2}t^2 \tag{5}$$

where $y_i(t)$ indicated the observed looking time for individual i on trial t, a_i was a

constant intercept value for infant i, and where b_{i1} and b_{i2} gave the linear and

quadratic slope terms for individual i, respectively. As with Dannemiller's (1984)

study, data were simulated under an exponentially decreasing proportional model

of the form: $L(t) = A_{\text{lower}} + (1.0 - A_{\text{lower}} - A_{\text{upper}}) \exp[-s(t-1)]$, where $L(t)$ is a

looking time response on trial t expressed as a proportional decrement such that

$0 \le L(t) \le 1$. A_{lower} and A_{upper} are the minimal and maximal looking times,

respectively, and s is a rate or slope term. Specific values of A_{lower}, A_{upper} and s

used to simulate data were randomly drawn from uniform distributions:

$A_{\text{lower}} \sim U(0.0, 0.2)$, $A_{\text{upper}} \sim U(0.0, 0.4)$ and $s \sim U(0.3, 1.0)$. Error terms of

systematically varying magnitude were included within each simulated

"observation", and generated values of $L(t)$ falling outside the range of $(0,1)$ were

disallowed.

These simulated data were used to generate parameter estimates for a_i, b_{i1}

and b_{i2} under Equation 5, and the parameter estimates used to compute the

predicted value of $y_i(t)$ for trial $t = 1$. An alternative to the 50%DC based

on Equation 5 was proposed such that habituation was deemed complete when

the fitted value of $\hat{y}_i(t = n) \le \dfrac{\hat{y}_i(t = 1)}{2}$ for $t = 1, 2, \ldots, n$, where n is a positive

integer greater than 1. This criterion was compared to the conventional 50%DC,

and as did Dannemiller (1984), they found that the number of trials used to

compute the 50%DC (e.g., 2, 3, or 4), along with the magnitude of error added to

observations, strongly influenced the validity of the 50%DC under their simulated conditions, in which the true values for the underlying exponential model were specified. Subjective comparisons were also made between the 50%DC and their polynomial regression criterion to determine which was the better predictor of the "true" trial corresponding to 50% relative habituation, as determined under each simulation with specified parameter values. They claimed that the average difference between the predicted and the true stopping points was smaller for their polynomial regression criterion than for the 50%DC, based on 5000 simulations, though no objective statistical measures were provided. Furthermore, given varying levels of (simulated) post-habituation recovery to a novel stimulus, their polynomial regression criterion was more sensitive than the 50%DC for detecting recovery after habituation. They did not substantiate this claim with any examination of real data, however.

While Ashmead and Davis' criterion showed the potential value in using all habituation responses rather than just the first and last few trials as the basis for characterizing individual performance, shortcomings are associated with this criterion (Thomas & Gilmore, in press). Specifically, while their simulated data were generated from a negative exponential model, the criterion is based on Equation 5, a polynomial regression model, the behavior of which can be strikingly different. More specifically, while a negative exponential model such as Equation 4 approaches some asymptotic minimal response level, as expected

under the typical habituation response, the polynomial model is unbounded or diverging. Furthermore, because the polynomial regression model is not necessarily monotonic with increasing trials, measurements of relative individual habituation based on the Ashmead and Davis could be theoretically difficult to interpret (i.e., predicted looking times increasing towards infinity).

The development and testing of an individual habituation model consistent with theoretical expectations about habituation, and the evaluation of such a model against real infant data would seem to be the next logical steps. Gilmore and Thomas (2002; Thomas & Gilmore, in press) proposed just such a model, but with a slightly different intent. Rather than assuming that all infants must be habituating (e.g., Colombo, 1993), the issue was evaluated as an empirical question. A regression model framework was proposed to describe habituation, based upon certain commonly accepted assumptions regarding the process as described below.

Proposed Model Framework

Under Thomas and Gilmore's (in press) model, habituation as measured by looking time on any given trial t was viewed as a random variable, $Y_i(t)$ for each infant i. Under their model, shown in Equation 6, any observed looking

time was interpreted as being a combination of some unknown true habituation function and an additive error term such that

$$Y_i(t) = h_i(t) + E_i(t) \text{ for all trials } t = 1, 2, \ldots, T_i, \qquad (6)$$

where $h_i(t)$ represents the unknown true function, T_i is the last habituation trial for infant i, and $E_i(t)$ is a continuous random error term symmetric around zero and independent for each trial, t. Since the expected value for $E_i(t)$ is zero, the expected value for the model $Y_i(t)$ is $h_i(t)$ for all trials t. Under this structure, one can estimate the unknown model parameters of $h_i(t)$ from the observed values $y_i(t)$, along with an estimate of the corresponding variance associated with E_i.

Certain additional properties or assumptions were imposed upon $h_i(t)$ in order to fashion a theoretically plausible model of the habituation response. The model assumed that $h_i(t)$ should have some maximum value at the onset $(t = 1)$, and that it should decrease over trials to some non-negative floor value, because observed looking times cannot be negative. Furthermore, it was assumed that with increasing trials, an infant's looking time response should approach its theoretical floor value increasingly gradually, exhibiting what Thomas and Gilmore called "soft landings".

Their proposed structure for $h_i(t)$, referred to here as $h_{3i}(t)$ for its three parameters, is shown here as Equation 7 :

$$h_{3i}(t) = \beta_i e^{-(\delta_i(t-1)^2)} + \alpha_i \tag{7}$$

where α_i, β_i, and δ_i are non-negative parameters estimated from individual data for each infant i. Figure 4 illustrates this model using parameter values taken from real data.
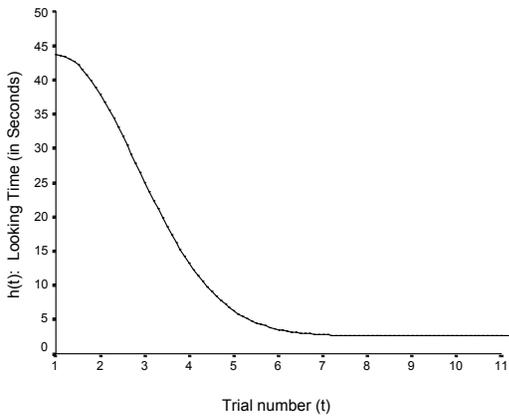
Figure 4: 3-Parameter Model of Habituation: $\alpha_i = 2.6$, $\beta_i = 41.06$, $\delta_i = 0.15$

Under this structure, α_i represents infant i's asymptotic floor, β_i represents the "depth" or range of the function, denoting the difference between the infant's minimum and maximum looking times, $h_{3i}(1) - h_{3i}(\infty) = \beta_i$, and δ_i represents the "slope" or rate of decline. Larger values of δ_i in the model indicate faster habituation over trials. Values of δ_i greater than about two or so cause $h_{3i}(t)$ to approximate a step function, essentially reaching its floor value, α_i, by $t = 2$.

The structure of $h_{3i}(t)$ differs from previous exponential models proposed in the literature. Dannemiller's (1984) model, shown here as Equation 4, does not include an independent range variable for each infant comparable to β_i, but uses a constant proportional scale for all infants. Inclusion of β_i makes it possible to fit observed data from infants with different response ranges. Under Dannemiller's model, used only for simulation, known values for the individual maximal and minimal responses were specified *a priori*, then used to rescale the response variable on a range from 1 (maximal response) to 0 (minimal response). In real world situations, however, these maximal and minimal values cannot be known beforehand, so the scaled model could not be directly fit to data. A preliminary model, directly fitting the maximal and minimal values, such as the one proposed by Thomas and Gilmore, would be required before Dannemiller's model could be used with actual data. After individual estimates of an infant's maximum and minimum response levels were fitted, then the entire sequence could be rescaled by expressing each fitted response as a proportion of the infant's range as Dannemiller did.

Ashmead and Davis' (1996) model, shown in Equation 5, also differs from $h_{3i}(t)$ in its use of polynomial fitting functions that are not, for polynomials of second degree or higher, monotonically declining with increasing trials, t. Consider Figure 5, which compares $h_{3i}(t)$ and Ashmead and Davis' model, using a
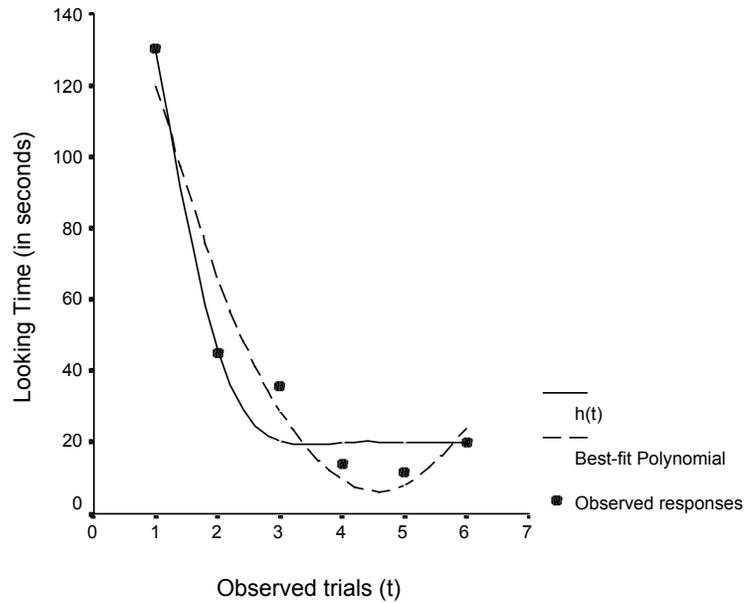
Figure 5: $h_{3i}(t)$ Compared to Ashmead and Davis' Polynomial Regression Model

sequence of observed looking times from the current study. Both models do a

decent job at explaining the observed variance in responses: $r^2 = .8959$ for

Ashmead and Davis' best fit polynomial, while $r^2 = .9429$ for $h_{3i}(t)$. Yet

Ashmead and Davis' model is less consistent with theoretical expectations about

habituation. Rather than predicting an asymptotic minimum with increasing

trials, as might be expected from a typical habituation response, the predicted

value of the best-fit polynomial is unbounded as trials increase.

With the goal of creating a theoretically plausible model that could be

used with real data, Thomas and Gilmore (in press) combined the general form of

Equation 6 with the structure of $h_{3i}(t)$ in Equation 7 to yield their proposed

habituation Model 3, so named for its three parameters, α_i, β_i, and δ_i:

$$Y_i(t) = \beta_i e^{-(\delta_i(t-1)^2)} + \alpha_i + E_i(t) \text{ for all trials } t = 1, 2, \ldots, T_i \qquad (8)$$

where T_i and $E_i(t)$ are as specified in Equation 6.

A special case for Equation 8 exists when there is no systematic decline in

the looking time response over trials. In this case, the model becomes a constant

function in which $h_i(t) = \alpha_i$, and any observed variance is attributed to error as

shown in Equation 9. This non-habituating model, one in which no systematic

habituation takes place, will be referred to henceforth as Model 1 for its single

parameter, α_i:

$$Y_i(t) = h_{1i}(t) + E_i(t) = \alpha_i + E_i(t) \qquad (9)$$

where α_i is a constant value for infant i, denoting the infant's mean looking time

over trials 1, 2, …T_i.

Thomas and Gilmore (in press) evaluated Models 3 and 1 in data from 35

infants deemed to have habituated under Ashmead and Davis' (1996) criterion.

20 infants demonstrated better agreement with Model 3 while 15 were better fit

by Model 1, meaning that more than a third of them may have achieved

Ashmead and Davis' criterion merely by chance. While Thomas and Gilmore's

approach demonstrated the potential for misclassifying infants who have achieved

habituation criteria without actually habituating, a number of additional questions might yet be addressed under this framework.

Habituators vs. Non-habituators: Multiple Types of Responders?

Analyses with Models 1 and 3 demonstrated how infants may achieve alternative habituation criterion merely through chance (e.g., Bogartz, 1965), particularly when there is large variance associated with their responses. Thus, Thomas and Gilmore's findings suggest that some infants who achieve the 50%DC may not actually be habituating. Yet most researchers who use infant-control paradigms are primarily interested in their subjects' post-habituation responses, under the belief that all infants who achieve the 50% decrement criterion have habituated to roughly equivalent states (e.g., Colombo, 1993), and are thus equally "prepared" to respond to the novel stimulus.

If, as Thomas and Gilmore (in press) suggested, some infants achieve this criterion without actually having habituated, then one could not reasonably expect such infants to respond to a novel stimulus in the same manner as infants who had truly habituated. More specifically, one might expect systematically more recovery of interest or dishabituation to a novel stimulus from infants classified under Model 3 (habituating) than infants classified under Model 1 (non-habituating), even if both groups had achieved the 50%DC.

Additional Habituation Models

Model 3, shown in Equation 8, was developed under an assumption of monotonic decline in the looking time response towards a fixed asymptotic minimum α_i as trials increase. For some infants, it may well be that $\alpha_i = 0$, in which case a simpler model emerges as shown in Equation 10:

$$Y_i(t) = h_{2i}(t) + E_i(t) = \beta_i\, e^{-(\delta_i\,(t-1)^2)} + E_i(t)$$
$$\text{for all trials } t = 1, 2, ..., T_i.$$

(10)

This model, henceforth referred to as Model 2, includes two parameters, β_i and δ_i, differing from Model 3 only in that its asymptotic minimum is fixed at zero. Figure 6 provides an illustration using values obtained from an infant in the current study:
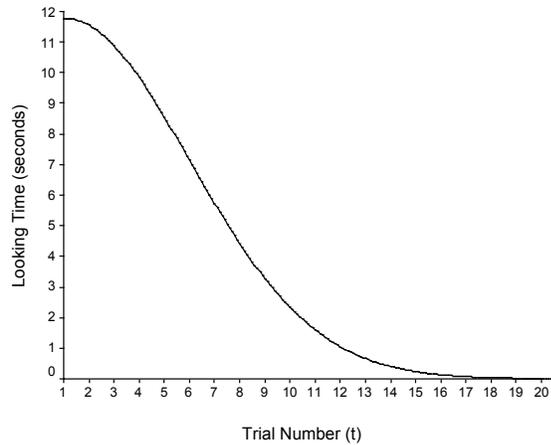
Figure 6: Model 2 With $\beta_i = 11.78$ and $\delta_i = 0.02$.

Dual process theory (e.g., Kaplan, Scheuneman, Jenkins & Hilliard, 1988; Kaplan, 1990) suggests that infants' looking time responses are driven by two opposing processes, an initial sensitization period in which looking times increase over trials followed by the typical habituation response. Model 3 can be generalized to accommodate an initial increase through the addition of a fourth parameter, μ_i, denoting the trial of maximum response. This is illustrated in Figure 7.
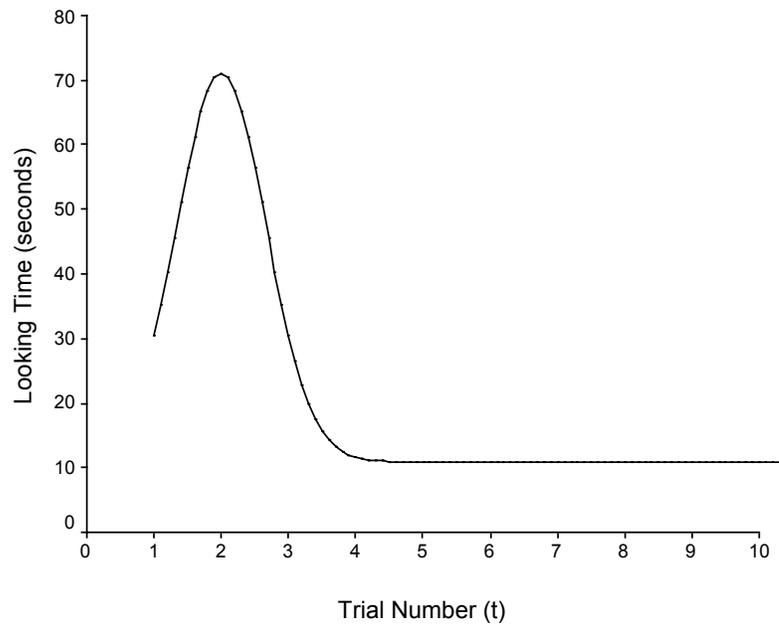
Figure 7: Model 4 With $\alpha_i = 10.9$, $\beta_i = 60.1$, $\delta_i = 1.12$, $\mu_i = 2$

This model takes the form: $h_{4i}(t) = \beta_i e^{-\delta_i(t-\mu_i)^2} + \alpha_i$ so that $Y_i(t)$ appears as in

Equation 11 (henceforth referred to as Model 4 for its four parameters):

$$Y_i(t) = \beta_i e^{-(\delta_i(t-\mu_i)^2)} + \alpha_i + E_i(t) \tag{11}$$

where μ_i is a positive integer denoting the trial at which maximum looking time

occurs for infant i. This parameter was constrained to an integer value for two

reasons: First, it was found that when μ_i was not so constrained, the

corresponding β_i estimates became artificially inflated to high values far beyond

those observed in real data. Additionally, since observable responses were only

available for discrete values of t, it seemed reasonable to similarly constrain μ_i as

well. In Model 3, μ_i is specified *a priori* at $\mu = 1$ for all infants, whereas under

Model 4 it would be estimated from each infants' observed responses.

Adding the fourth parameter might not necessarily yield more useful

results from a practical standpoint, however. If one's motivation for using a

modeling approach is to better distinguish between habituating and non-

habituating infants, the additional parameter may not necessarily lead to

significantly different results, since it is the response time on the latter trials

immediately before the novel stimulus, rather than the first few trials, that would

primarily distinguish habituating and non-habituating responders (i.e., did the

infant's responses approach some asymptotic minimal level as the number of

trials increased?) The law of parsimony, one of the guiding principles of scientific

research, suggests that when all other things are equal, the simpler model is the more desirable. Therefore, one could compare the effectiveness of Models 3 and 4 to determine whether the inclusion of the non-monotonic Model 4 increased one's ability to distinguish habituating from non-habituating responses. If it did not, then parsimony would favor using only Model 3.

Using an Infant's Response Model to Predict Future Behavior

One goal of formal modeling is to identify and quantify systematic relationships among variables of interest. Moreover, the extent to which such relationships are consistent across data sets or over time provides an index of how generally applicable the model might be. Ideally, one might hope for broadly generalizeable models so that predictions can be made about future outcomes.

Specifically, if model-based classifications truly capture an underlying cognitive process within infants (i.e., differentiating between non-habituating and habituating responders) then one might well ask whether this behavior generalizes to new stimuli, or how consistent the behavior remains over time. One might expect different models to depict different looking behaviors for different visual stimuli. For example, dual process theory implies that stimulus characteristics drive the habituation process (Kaplan & Werner, 1986; Kaplan &

Werner, 1987). With stimuli of a similar nature, however, model parameter estimates generated during the habituation phase might well predict behavior during the test phase if the model is capturing some systematic process within the infant. Specifically, if h_i^* is an infant's best-fit habituation phase model, then one might expect that some model $h_{i,ph}^*$ characterizes post-habituation responses where h_i^* and $h_{i,ph}^*$ have some unknown functional relationship.

Similarly, one might ask whether patterns of habituation behavior observed on one day predict future habituation a week later, or two months later. For example, does model classification remain stable for infants across multiple visits separated by a week or more? How reliable are the specific fitted parameter values under the best-fit models across separate visits? Examining infants' habituation behavior across multiple visits separated by intervals of varying lengths would provide the means for examining such questions of stability and reliability. More specifically, 4-month old infants could be seen twice over a one-week interval, then twice again at 6 months. Each infant's looking times could be fit to Models 1-4 to observe the consistency of infant habituation responses, and how reliable individual parameter estimates remain over the given intervals. Which parameters show the greatest and least stability? Naturally, one would expect variation in looking behavior over different occasions due to transient conditions such as infant fatigue, irritability, and hunger. Nevertheless, since previous research has found modest levels of reliability for

habituation measures based upon conventional and Ashmead and Davis stopping criteria (e.g., Lavoie & Desrochers, 2002), one might well ask how reliable are these model-based measures over similar time intervals, and how such reliability statistics compare to those generated by conventional criteria.

Current Study

The current study, building on earlier work, gets its conceptual underpinnings from Gilmore and Thomas (2002) and Thomas and Gilmore (in press) and uses the general methodology of Gilmore and Rettke (2003), in which infants viewed a repeating series of computer-generated visual stimuli on a large television monitor while their responses were judged by an unseen experimenter. The current study uses a similar infant control procedure, employing the conventional 50%DC so that direct comparisons could be made between model-based and conventional measures. A longitudinal design was used to address issues of stability of model-based classifications across two visits separated by a weeklong interval. Between subjects comparisons were also made for three categories of visual stimulus to search for potential effects of stimulus type on habituation responses.

In addition, the current study directly compared the efficacy of model-based vs. conventional measures to detect significant dishabituation responses to

novel stimuli. This was achieved through the use of matched pairs of visual stimuli (i.e., habituating and novel stimuli) that were already known to be visibly distinguishable among infants of the targeted age group. Use of such matched stimuli eliminated the possibility that non-significant dishabituation responses might be attributable to infants' inabilities to distinguish between the familiarized and novel stimuli, while preserving maximum possible similarity between the repeating and novel stimuli shown to each infant. Direct comparisons were made between conventional procedures for measuring post-habituation recovery such as matched pair t-tests or repeated measures ANOVAs (e.g., Kagan, Linn, Mount, Reznick & Hiatt, 1979; Hunter, Ames, & Koopman, 1983; Kirkham, Slemmer & Johnson, 2002) and the proposed model-based measures to assess which was better at detecting post-habituation recovery responses.

Chapter 2

Methods and Analytical Strategies

The current chapter is divided into two major themes: methods and

analyses. The first provides descriptions of the participating subjects, the design

and stimulus groups used, the methods and the equipment used to collect the

data. The methodology was adapted from Gilmore and Rettke's (2003) motion

perception studies using an infant control habituation paradigm. The second

major theme of the chapter outlines the analytical strategies used, such as model

fitting, selection strategies, and bootstrap-based inferential statistics for groups

and individuals.

Participants

Fifty-six 4-6 month old (107-200 days; twenty-eight female) healthy

infants participated. Forty-five (twenty-one female) were seen twice in the lab,

with an interval of approximately one week between their first and second visits.

The other eleven infants were seen only once, as their parents were unable to

schedule a second lab visit. Data from one of these eleven children were

discarded because the infant did not complete the data collection procedure. In all, data from $n = 55$ infants collected over $v = 100$ visits are included in the analyses of the current study. All but two infants were within four weeks of normal gestational age at birth, according to parent report. The two premature infants were born approximately six weeks early, and participated when they were 5.5-months old rather than 4. Infants were recruited by telephone using information obtained from public birth records. Parents provided informed consent before any infants participated.

Design

Each infant was randomly assigned to one of the three stimulus conditions: Condition 1: Checkerboards, Condition 2: Optic Flow, and Condition 3: Faces; these conditions are detailed below. Although not a precisely balanced design (i.e., cell sizes were not precisely equal), conditions were roughly counter-balanced for sex. Infants within each condition were randomly assigned to one of two sub-groups to counter-balance the order of stimulus presentation, as indicated in Table 1.

Table 1: Visual Stimulus Conditions by Visit

Stimulus Condition	Visit 1	Visit 2
	Habituating Stimulus	Habituating Stimulus
1. Checkerboards	1a. horizontal	1a. vertical
	1b. vertical	1b. horizontal
2. Optic Flow	2a. forward right	2a. forward left
	2b. forward left	2b. forward right
3. Faces	3a. infant 1	3a. infant 2
	3b. infant 2	3b. infant 1

Visual Stimuli

Each of the visual stimulus conditions utilized a distinct matched pair of animated visual stimuli. Each matched pair consisted of two similar but perceptually distinguishable images. Similar images were selected to minimize the potential that any observed "dishabituation" or recovery response was attributable to the fact that one of the two images was inherently more engaging to the infants. As described below, all three stimulus conditions reflected previous studies of habituation and/or perceptual development. Infants within the stimulus conditions were counter-balanced such that one half of the group

saw one image from the pair as the habituating stimulus and the other image as the novel stimulus. All infants saw the same pair of stimuli in reverse order on their second visits.

Stimulus Condition 1: Checkerboards

Infants in this condition saw a pair of scrolling checkerboard patterns generated using Matlab's Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). The pair of images differed with respect to the orientation of their animation, with one image scrolling along its horizontal axis while the other scrolled on its vertical axis. Infants in this group viewed one image as the repeating, or habituating stimulus and the other as the novel stimulus. A static version of these images is shown in Figure 8:
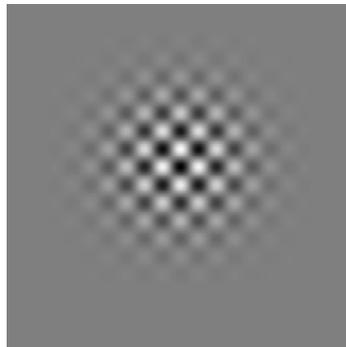
Figure 8: Checkerboard Visual Condition

Checkerboard patterns have commonly been employed in habituation studies (e.g., Kaplan & Werner, 1986; Kaplan & Werner, 1987; Kaplan, Scheuneman, Jenkins, & Hilliard, 1988), and have been recommended for the ease with which they can be manipulated to control for various stimulus characteristics, such as color or complexity.

The total display region was 40° horizontal (H) by 30° vertical (V) of visual angle (a standard 32 inch television monitor), although the checkerboard images appeared in a central rectangular region of the viewscreen in an area that was 20° horizontal (H) by 20° vertical (V), or roughly 32 cm square (12.5 inches) at a viewing distance of 90 cm.

In between each presentation of the checkerboard images, a short film clip of roughly one second duration was randomly selected from among ten unique clips and shown to re-orient the infant to the center of the screen. These clips included excerpts from children's television such as Sesame Street, Barney and Dr. Seuss, as well as images of other infants' faces. These short clips included sound.

Stimulus Condition 2: Optic Flow

Infants saw images portraying repeated forward-backward motion through a room filled with columns (as shown in Figure 9). These patterns of motion,

Figure 9: Optic Flow Stimulus Condition

called optic flow, have been used to study infants' perception of directional

motion (e.g., Gilmore & Rettke, 2003) within both habituation and forced

preferential looking paradigms. Gilmore, Baker and Grobman (in press) provide

a detailed account of the movement feature of these images.

The total display region was 40° horizontal (H) by 30° vertical (V) of

visual angle, although the optic flow images appeared in a central rectangular

region of the viewscreen in an area that was 15° horizontal (H) by 30° vertical

(V), or roughly 24 cm wide by 48 cm high at a distance from the viewer of 90

cm. The pair of animated images in this condition simulated motion at headings

of either forward left 22.5 degrees or forward right 22.5 degrees. These optic flow

images were generated using Matlab's Psychophysics Toolbox. The animated

stimuli simulate forward/backward motion in a forward left vs. forward right

heading from this initial starting point. Movement in both of these images

continued in a repeating forward/backward sequence in which the direction

reversed every 0.83 seconds.

As with the infants in the checkerboard condition, a random film clip was

shown between each trial or presentation to re-orient the infant towards the

center of the screen.

Stimulus Condition 3: Faces

Infants saw animated film clips showing one of two unique smiling baby

faces as shown in Figure 10, which depicts the first frame of both face conditions:

Figure 10: Infant Face Visual Condition

The total display region was 40º horizontal (H) by 30º vertical (V) of visual

angle, although infant face images appeared in a central rectangular region of the

view screen in an area that was 16º horizontal (H) by 12º vertical (V), or roughly

25 cm wide by 19 cm high at a viewing distance of 90 cm. Twenty-five frames

from each clip were shown in a forward-backward loop (at 30 frames per second),

continuously repeating itself every 1.63 seconds for the duration of the trial.

These images had previously been used as inter-trial re-orienting stimuli in

studies of perceptual development (e.g., Gilmore & Rettke, 2003) and were

known to be highly engaging to infants. These face clips were digitized excerpts

taken from videos collected in the lab, along with written authorization for the

clips to be used for research purposes. These short clips did not include sound.

Additionally, in between each presentation of the face images, a randomly

generated audio signal of one to five sequential tones, roughly one second in total

duration, was emitted to re-capture attention and re-orient the infant's gaze

towards the television monitor. This audio signal was not accompanied by a

visual clip, as were the two other stimulus conditions.

Procedure

Data were collected at the Infant Perception and Cognition Laboratory at the Penn State Child Study Center. Infants were placed in a car seat facing a large television monitor on which the images were shown. Distance between the monitor and the infants' faces was fixed at 90 cm. A video camera was positioned above the monitor to capture the direction of the infant's gaze without permitting the experimenter to see the actual image being viewed by the infant. The timing for each trial began when the infant looked at the image generated on the monitor and ended when the infant looked away, as judged by the experimenter. These events were recorded by pressing computer keystrokes. As per convention under the infant control paradigm (Horowitz, Paden, Bhana & Self, 1972b), trials had no maximum duration, ending only when the infant looked away from the image for at least two seconds. If the infant briefly looked away but looked back again before two seconds had elapsed, a new intra-trial "fixation" was noted and timing for that trial continued. The number of fixations and the duration of each were recorded. The sum of fixation durations determined the trial duration, $y_i(t)$.

Independent reliability assessments were performed on experimenter judgments of trial durations $y_i(t)$ using the videotapes from a random selection of 20% of the infant lab visits. These assessments revealed an average Pearson's

$r = .977$ (range $= .894$ to $.999$) between the experimenter's and the independent observer's looking time assessments, computed individually for each infant visit.

An infant's baseline response was defined as $\bar{y}_{bi} = \dfrac{\sum_{t=1}^{3} y_i(t)}{3}$. Beginning with the sixth trial, a running average $\bar{y}_{ci}$ was computed after each trial, based on the current and previous two trial durations. Habituation trials continued until $\bar{y}_{ci} \leq \dfrac{\bar{y}_{bi}}{2}$ (i.e., the 50%DC was achieved) or until a maximum of 15 habituation trials were completed. At this point, the post-habituation phase began, during which infants saw alternating trials of the familiar and novel and stimuli for a fixed total of six more trials. Post-habituation trials were recorded and terminated in the same fashion as the habituation trials. In total, each infant saw between twelve and twenty-one trials on a single experimental visit.

Apparatus

Visual stimuli were displayed on a Sony 32" video monitor. Images were generated on an Apple Macintosh G3 computer. Infants rested on a car seat during testing. Infants' responses (eye and face movements) as they viewed the visual stimuli were captured by a video camera mounted above the video display and conveyed to a second smaller video monitor visible to the experimenter. The walls of the small testing room and all equipment visible to the infant were

covered by dark cloth and the interior lights were dimmed to minimize potential distraction to the infants.

Analytical Strategies

Although there were advantages to applying a model-based approach to the study of infant habituation behavior, they were not without cost. Use of nonlinear iterative models is analytically more complex than linear regression models, for which closed form solutions are available. Furthermore, many of the commonly recognized sample statistics typically used to draw inferences about their underlying populations simply did not apply, given the unusual characteristics of the habituation data, both at the individual and group levels.

Thomas and Gilmore (in press) note that many group-based analyses become problematic when applied to habituation data collected from individuals whose responses are best described by differing probability distributions. Yet the notion of independent, identically distributed (*iid*) data is a lynchpin of many conventional statistical approaches. The matched pairs *t*-test, for example, is often used to detect systematic differences in average infant looking time responses to familiar and novel stimuli. Yet this analysis assumes that the expected difference (in seconds) between a novel response and a familiar response is constant for all individuals, attributing any other variation to random error.

Failure to recognize that all infants do not share the same expected maximal and minimal response levels could result in larger estimates of error variability, potentially reducing the power of the test.

The use of some ordinary sample statistics was problematic within the current dataset as well. For example, Pearson's r, the conventional measure of correlation between two continuous variables, considers the variability across individuals of the measured quantities, but does not consider the uncertainty associated with the individual measurements. Yet in the current study, considerable variation in error magnitude was found among individuals for certain estimated quantities. Among habituating infants, for example, the standard errors associated with individual estimates of the slope parameter $\hat{\delta}_i$ ranged from 0.005 to 25.425, with a median of 0.183 and a mean of 5.242 (estimation techniques for δ_i and all other model parameters are detailed in the following section). Estimates of Pearson's r between $\hat{\delta}_{i1}$ and $\hat{\delta}_{i2}$, estimated for an infant at visits one and two, respectively, do not consider individual standard errors, and may therefore be misleading. Failure to consider such variability of measurements, both within and across individuals, necessitated the use of alternative approaches for several underlying research questions.

The following sections discuss the analytical strategies used to deal with these issues, including the model estimation and selection strategies, tools for estimating the standard errors associated with model parameter estimates and

fitted values (i.e., their precision), as well as some of the problems encountered and decisions made along the way.

Estimation of Nonlinear Regression Models

Infants' sequences of habituation phase looking time responses were fit to Models 1, 2, 3 and 4 using algorithms written by the author for the Matlab Curve Fitting Toolbox (see Appendix B, for example, for the algorithm used to estimate Model 3). Parameter estimation was achieved via Trust-Region optimization (e.g., Branch, Coleman & Li, 1999; Seber & Wild, 1989, p. 603), a method that permitted constraints on the fitted parameter values with $\hat{\beta}_i$, $\hat{\delta}_i$, and $\hat{\alpha}_i \geq 0$, and $\hat{\mu}_i = 1, 2, 3$ or 4.

The Trust-Region Method is an iterative non-linear least squares estimation technique. Given a series of constraints and starting values for the parameters in the model, a set of squared residuals based on the differences between the observed values and the starting values was computed. Model parameter estimates were updated after each iteration so as to minimize the sum of these squared residuals using information from the Jacobian matrix, a square matrix of partial derivatives equal in dimension to the number of parameters under the model. Iterations continued until either the specified maximum number of iterations had occurred, or until the differences in the estimated

parameter values between iterations decreased to a specified minimum tolerance level. The current study used Matlab's default fitting options for the Trust-Region algorithm, which included a maximum parameter adjustment between iterations of 0.1, a maximum number of iterations of 400, and a termination tolerance between iterations for parameter estimates of 0.000006.

These default fitting options produced nearly trouble-free results with the possible exception of the $\hat{\delta}_i$ estimates. In the course of the study it was observed that fitted values of $\hat{\delta}_i > 2$ produced essentially a step function with, for example, $h_{i3}(t) = \alpha_i$ by approximately $t = 2$. Any fitted $\hat{\delta}_i > 2$ produced residuals that were essentially equivalent, creating a tendency for the fitted $\hat{\delta}_i$ values to explode during the estimation process. Larger $\hat{\delta}_i$ values produced no ill effects in terms of adequacy of model fit, but the resulting bootstrap-estimated standard errors (detailed below) associated with these fitted $\hat{\delta}_i$ values tended to be artificially high.

Starting values for the parameter estimates were generated in the current study using a direct search method (e.g., Everitt, 1987, p. 23-24) by specifying a range of values for each parameter under the model, and generating a finite number of point estimates uniformly distributed within that range. Summed, squared residuals were computed for each set of observed data using all possible combinations of point estimates, and the single combination of parameter point estimates that gave the lowest residual value (with respect to the observed

sample) was used as the set of starting points under the fitting algorithm. The range of possible starting values for α_i, β_i, and δ_i were bounded below at zero, and above by a maximum value computed as a function of the observed data, or from a fixed maximum. Consider the following observed looking time responses in Table 2 from an infant visit within the current study:

Table 2: Decreasing Looking Time Responses Over Repeating Trials

Habituation Trial Number	1	2	3	4	5	6
Looking Time Response (seconds)	12.35	5.71	4.47	1.67	1.60	2.44

For α_i, a range between 0 and the minimum observed response plus six was used. In Table 2, this corresponds to $1.60 + 6 = 7.60$. For β_i, a range between zero and the maximum observed response plus twenty was used (or $12.35 + 20 = 32.35$ in Table 2). For δ_i, a fixed range between 0 and 10 was always used. Each range of possible starting points was divided into 20-35 intervals for each parameter, depending on the model, such that a maximum number of residual computations under the direct search was no more than about 15,000. This number seemed to produce starting values that were essentially the same as when smaller increments were used, but could be done within two to three seconds.

This strategy was used to generate a set of parameter estimates for Models

2 through 4 and an associated estimate of error variance for each, based on the

residuals between observed and fitted responses under the fitted parameter

estimates. For Model 4, however, the data were fit to Model 3 four times in

succession, each time with μ_i fixed at 1, 2, 3 or 4 respectively. Of these four

values of μ_i, the fit with the lowest associated error variance was selected. For

the non-habituating Model 1, the simple arithmetic mean estimated $\hat{\alpha}_i$, and the

observed sample variance estimated $\hat{\sigma}^2_{h^*_i}$, the error variance under the model.

Classification of Looking Time Sequences to Models

Once the parameter estimates and corresponding error variances were

computed under each model, each sequence of responses was classified as

"belonging to" the model with the lowest corresponding BIC value (Ramsey &

Schafer, 1997, p. 343), as shown in Equation 12:

$$BIC_{ik} = T_i \log(\hat{\sigma}^2_{ik}) + k \log(T_i) \tag{12}$$

where BIC_{ik} denotes the value for infant i under the model and T_i denotes the

number of habituation trials observed for infant i; k denotes both the number of

estimated model parameters in the model and the model name. $\hat{\sigma}^2_{ik}$ denotes the

observed error variance for infant i under model k. The model k with the lowest

associated BIC_{ik} is selected for infant i. Use of BIC, rather than merely

classifying each sequence of responses under the model with the lowest overall

error variance, helped to reduce the possibility of "over-fitting" the data by

placing a penalty on models with larger numbers of parameters. BIC, therefore,

favors the best combination of few parameters and low error variance.

There was always only one best-fit model with minimum BIC.

Nevertheless, competing models occasionally produced very similar BIC values,

making the classification less certain. Table 3, for example, shows the

habituation phase looking time responses for an infant in the current study, along

with the corresponding BIC and fitted parameter estimates for the four models:

Table 3: BIC and Parameter Estimates for a Given Set of Observations

Habituation Trial Number	1	2	3	4	5	6
Looking Time Response (in seconds)	41.36	12.67	19.05	7.16	2.04	4.30

	BIC	$\hat{\alpha}_i$	$\hat{\beta}_i$	$\hat{\delta}_i$	$\hat{\mu}_i$
Model 1	33.9239	14.4304	--	--	--
Model 2	29.7426	--	32.2124	0.1940	--
Model 3	29.6587	8.1083	33.2401	1.9607	--
Model 4	33.8832	8.1083	33.2401	1.9607	1

Although this infant's responses produced the lowest BIC value for Model 3, Model 2 produced a value very nearly as low. Additionally, it may be seen that Models 3 and 4 produced equal parameter estimates, though the BIC value associated with Model 4 is higher. This reflects the penalty factor under BIC, which favors models with fewer numbers of estimated parameters.

Bootstrap Estimation of Standard Errors

Once the best-fit model and corresponding model parameters were estimated, standard errors were computed by bootstrapping the residuals (Efron & Tibshirani, 1993; pp. 111-113). Bootstrap-estimated standard errors for fitted parameters such as $\hat{\sigma}_{\hat{\delta}_i}$ provide an index of precision for the estimated parameter values under the best-fit model for each observed sequence of infant responses. These estimates were used in other computations as well.

Table 4 shows the strategy used in the current study to bootstrap the residuals, as described by Efron and Tibshirani (1993; pp. 111-113):

Table 4: Standard Errors of Regression Parameters by Bootstrapping the
Residuals

1. Find the regression model parameter estimates for the coordinates t and
    $y_i(t)$, $t=1, 2, ..., T_i$.

2. Compute the vector of residuals, r, where $r = y_i(t) - h_i^*(t)$ for $t=1, 2, ..., T_i$.

3. Draw B samples (with replacement) from the residual vector r of sample
    size T_i to create B boot residual vectors r_{boot}. B≥250.

4. For each r_{boot}, compute the vector sum $y_i(t) + r_{boot}$ to create B new boot
    samples y_{boot}.

5. Compute the regression model parameter estimates for the coordinates t
    and $y_{boot}(t)$ for $t= 1, 2,..., T_i$ for all y_{boot} vectors. Store the resulting
    parameter estimates in a B x k matrix, where k denotes the number of
    parameters in the regression model. Each row in the matrix corresponds
    to a boot sample y_{boot}.

6. Compute the sample standard deviation for each column of the B x k
    matrix, and take this value as the standard error for that model
    parameter.

Under this stragey, bootstrap standard errors for parameter estimates could be
computed relatively quickly for Models 1 though 3. Each bootstrap sample was
regressed using the nonlinear regression estimation technique described above, so
that standard errors for a specified model and dataset were completed within a
few minutes using B = 250. Bootstrap standard error computations under Model
4 typically required about an hour using B = 250.

Bootstrap Hypothesis Testing of Group   Means

Hypothesis tests comparing the means of two or more groups representing

differing levels of some categorical independent variable (e.g., habituating vs.

non-habituating infants, or among infants shown different categories of visual

stimuli) are typically accomplished through the use of conventional t-tests and

oneway analyses of variance. As a means of avoiding potential violations of the

assumptions underlying standard hypothesis tests, group mean comparisons in

the current study were accomplished through the use of bootstrap hypothesis

tests (Efron & Tibshirani, 1993, pp. 220-226). These tests use an iterative re-

sampling technique to directly estimate the sampling distribution and

corresponding p-level for the statistic of interest. In the two group case, Efron

and Tibshirani's Algorithm 16.2 (1993, p. 224) was used. In the case of 3 or

more groups, a modification was used as shown in Table 5:

Table 5: Bootstrap Oneway ANOVA Algorithm

1. Compute the F-statistic from the observed groups.
2. Adjust the observed values within each group to that the condition of equal group means under the null hypothesis holds true.
3. Draw B samples (with replacement) from each of these adjusted groups, where the size of each boot sample group equals the size of the observed group. $B \geq 250$.
4. Compute the value of the F-statistic for each boot sample.
5. Estimate the bootstrap achieved significance level (p_{boot}) by tallying the proportion of bootstrap samples whose corresponding F-statistics were at least as large as the observed statistic.

In the event that a bootstrapped oneway ANOVA yielded a significant main effect, a multiple comparisons procedure examined the individual cell means. This was done by constructing bootstrap-generated confidence intervals around the mean differences for each pairwise comparison with a familywise error rate of $p = .05$. Table 6 describes the algorithm used:

Table 6: Bootstrapped Multiple Comparisons Procedure

1. Compute the means and all possible pairwise mean differences from the observed groups. Let, for example, $\bar{x}_1$, $\bar{x}_2$, and $\bar{x}_3$ denote the observed sample means of groups 1, 2, and 3, respectively, and d_{12}, d_{13}, and d_{23} denote the corresponding group mean differences.

2. Sample with replacement from each of the observed groups to construct a boot group (e.g., x_{1_b}, x_{2_b}, and x_{3_b}), where the size of each boot group equals the size of the observed group. Compute the boot group means (e.g., $\bar{x}_{1_b}$, $\bar{x}_{2_b}$, and $\bar{x}_{3_b}$) and all possible boot group mean differences (e.g., d_{12_b}, d_{13_b}, and d_{23_b}). Store each computed boot group mean difference in a column vector.

3. Repeat this process for B $= 3600$ bootstrap repetitions.

4. Rank order the boot group mean differences within each column from lowest to highest (e.g., one column for d_{12_b}, one for d_{13_b}, and so forth). Each column estimates the sampling distribution for that pairwise difference.

5. Compute the upper and lower endpoints of the confidence intervals of each pairwise mean difference by dividing the familywise error rate ($p = .05$) evenly among the two distributional tails for all pairwise mean differences. For example, with a familywise error rate of $p=.05$, B=3600 bootstrap repetitions and 3 pairwise differences, the endpoints would be the lowest and highest $\dfrac{3600 * (.05)}{2 * 3} = 30$ values within each column. These values represent the upper and lower endpoints of the estimated confidence intervals for each pairwise mean difference.

6. Only those pairwise mean differences whose confidence intervals do not include 0 are concluded to differ significantly.

Measuring the Post-Habituation Recovery to a Novel Stimulus

One of the critical questions of the current study was whether model-based approaches provide any practical advantage over conventional methods for detecting a recovery, or "dishabituation" response to a novel stimulus after habituation. Also of interest was whether infants classified under different models exhibited discernible differences in their dishabituation responses. Specifically, it was hypothesized that habituating infants would demonstrate greater dishabituation to a novel stimulus than non-habituating infants since they would presumably be more inclined to perceive the novelty of the new stimulus. As mentioned earlier, the matched pairs *t*-test is the conventional analysis used to compare the group mean response differences between novel and familiar stimuli. Yet this test presupposes that all infants share a constant fixed difference, attributing any observed variation to random error.

Alternatively, the current study measured infants' dishabituation responses relative to their individual minimal and maximal response levels, estimated during the habituation phase. Assuming that infants ought exhibit no systematic dishabituation if novelty was not perceived, dishabituation responses were expressed as z_i-scores reflecting the magnitude of departure from the expected value under each infant's initial model, considering both the error variance under the best-fit model and the error variance associated with the

predicted looking time. Dishabituation z_i-scores were computed for each infant

under Equation 13:

$$z_i = \frac{y_i(T_i + 2) - h_i^*(T_i + 2)}{\sqrt{\hat{\sigma}^2_{h_i^*} + \hat{\sigma}^2_{h_i^*(T_i+2)} - 2\hat{\rho}(\hat{\sigma}_{h_i^*})(\hat{\sigma}_{h_i^*(T_i+2)})}} \tag{13}$$

where $y_i(T_i+2)$ denotes the observed looking time response for the infant's

second post-habituation trial (i.e., the first presentation of the novel stimulus)

and $h_i^*(T_i + 2)$ denotes the predicted looking time response under the infant's

best-fit model h_i^*. $\hat{\sigma}^2_{h_i^*}$ denotes the estimated error variance during the habituation

phase under the infant's best-fit model and $\hat{\sigma}^2_{h_i^*(T_i+2)}$ denotes the bootstrap-

estimated variance associated with the predicted value at trial T_i+2 under the

infant's best fit model. $\hat{\rho}$ denotes the estimated correlation between $y_i(T_i+2)$

and $h_i^*(T_i + 2)$ for all habituating responders or non-habituating responders,

computed separately.

Stability and Reliability of Model-based Measures

If the habituation behavior towards an initial stimulus successfully

predicted an infant's reaction to a similar but perceptually distinguishable novel

stimulus presented immediately thereafter, this would suggest that the underlying

cognitive/perceptual mechanisms driving both reactions were the same, or at

least related. This seems a reasonable assumption, given that the stimuli in

question are perceptually similar and presented to the infant within a very short time frame. Nevertheless, in order to infer any kind of longer term stability to the perceptual mechanisms, it would be desirable to show that infants tended to react similarly on separate occasions, separated by some amount of time. Failure to find such stability could suggest either that the measures are not sufficiently reliable, or that that the underlying cognitive/perceptual mechanisms themselves are subject to variability over time.

The current study addresses this question specifically by examining both the stability of model-based classifications (i.e., does the same model describe an infant's looking time responses on both visits?), as well as the reliability of individual model parameters for infants tested twice over a one-week interval. Stability was examined through the use of contingency table analyses, testing the null hypothesis of no relationship between model-based classifications for infants on their first and second visits. Reliability of the individual parameter estimates for each infant i was examined in two ways. Pearson's r indexed the correspondence between parameter estimates across time without consideration of the precision of the fitted values, as indexed by the magnitudes of the standard error of the estimates. A $z_{\hat{\theta}_i}$ score was also computed for each pair of estimates, using the computation shown in Equation 14:

$$z_{\hat{\theta}_i} = \frac{\hat{\theta}_{i1} - \hat{\theta}_{i2}}{\sqrt{\hat{\sigma}^2_{\hat{\theta}_{i1}} + \hat{\sigma}^2_{\hat{\theta}_{i2}} - 2\hat{\rho}(\hat{\sigma}_{\hat{\theta}_{i1}})(\hat{\sigma}_{\hat{\theta}_{i2}})}} \qquad (14)$$

where $\hat{\theta}_{i1}$ and $\hat{\theta}_{i2}$ represent the estimated values for a generic model parameter θ (e.g., β_i, α_i or δ_i, as the case may be) for infant i at times one and two, respectively. $\hat{\sigma}_{\hat{\theta}_{i1}}$ denotes the bootstrap estimated standard error of the fitted parameter $\hat{\theta}_{i1}$. $\hat{\rho}$ denotes the observed correlation between $\hat{\theta}_{i1}$ and $\hat{\theta}_{i2}$, estimated over all infants consistently classified (i.e., habituating) over both visits.

<center>**Chapter 3**</center>

<center>**Results**</center>

This chapter is organized into separate sections, each including graphical and quantitative information specifically related to one of the research questions or issues posed in the preceding chapters. The first section provides initial descriptions of the magnitude of individual differences, both within and across conditions. Figures and tables indicate the number of infant visits (v) under each of the visual stimulus conditions, the number of visits classified under each model type according to BIC, and how the behavior under the various model types differed. Subsequent sections examine group differences, post-habituation recovery responses to perceptually similar but distinguishable (novel) stimuli, predictive validity and reliability of habituation behavior for individual infants across visits.

Individual Differences in Habituation Response Behavior

A total of $n = 55$ infants were tested for a total of 100 visits (45 infants x 2 visits + 10 infants x 1 visit = $v = 100$). Each infant's looking time responses

during the habituation phase were individually fit under Models 1 through 4, and classified under the one with the lowest associated BIC value. Table 7 provides a summary of these model fits within each visual stimulus condition. Evidence will be presented later on suggesting that each infant's visit was essentially an independent observation (i.e., a null hypothesis of independence of model classification across visits for infants seen twice could not be rejected). Subsequently, Table 7 collapses across multiple visits for infants tested twice, describing the model classifications from $v = 100$ visits rather than $n = 55$ infants. Estimated means and medians for the various parameters under each model are also shown, as well as the standard errors of the mean for each parameter. Standard errors of the mean are calculated for each estimated parameter as shown in Equation 15:

$$SEM(\hat{\theta}_i) = \sqrt{\sigma_{\hat{\theta}_i}^2} = \sqrt{\frac{\sum_{i=1}^{v}(\hat{\sigma}_{\hat{\theta}_i})^2}{v^2}} \tag{15}$$

where $\hat{\theta}_i$ represents a generic parameter (e.g., α, β or δ) estimated for infant i, $\hat{\sigma}_{\hat{\theta}_i}$ denotes the bootstrap estimated standard error of $\hat{\theta}_i$ and v denotes the number of infant visits being considered. Table 7 shows both the means and medians for the various parameter estimates within each model and stimulus condition, as many of the distributions of estimated parameters were non-symmetric. Standard errors of the means also show the degree of variability of estimated parameters across the models and stimulus conditions.

Table 7: Mean and Median Parameter Estimates by Visual Stimulus Condition

Visual Stimulus	Model Number	Infant Visits v	Mean Parameter Estimate (with SEM) and Median Parameter Estimate		
			$\hat{\alpha}$	$\hat{\beta}$	$\hat{\delta}$
Checkers	1	7	4.49 (0.52)	--	--
			4.21		
	2	10	--	9.55 (0.65)	0.12 (0.01)
				6.86	0.07
	3	13	4.53 (0.32)	19.72 (0.84)	12.76 (3.18)
			4.36	14.24	1.33
	4	2	10.82 (2.17)	40.93 (7.48)	38.15 (11.47)
			10.82	40.93	38.15
Optic Flow	1	9	7.48 (0.86)	--	--
			6.23		
	2	11	--	28.31 (1.41)	0.07 (0.02)
				25.11	0.06
	3	15	6.49 (0.44)	34.02 (1.06)	8.87 (3.03)
			6.05	34.15	1.73
	4	1	4.01 (2.96)	62.65 (6.47)	1.67 (3.28)
			4.01	62.65	1.67
Faces	1	9	16.39 (1.66)	--	--
			12.86		
	2	6	--	25.16 (1.99)	0.03 (.07)
				22.67	0.03
	3	12	9.29 (0.65)	81.23 (1.44)	9.91 (3.29)
			8.33	42.76	2.03
	4	5	10.53 (1.89)	57.08 (4.67)	9.14 (5.21)
			9.75	49.12	1.79

Note: Standard errors of the mean (in parentheses) computed as in Equation 15

Qualitatively different types of habituation behavior were observed within each of the three visual stimulus conditions, as implied by the fact that all four models best described at least some infants' responses. Model 3 best captured the responses for the greatest number of infants within each stimulus condition, while Model 4 did the poorest. For the eight infant visits best described by Model 4, seven of them showed a maximum estimated looking time on $\hat{\mu} = 2$, while one showed a maximum on $\hat{\mu} = 4$.

Within each of the three conditions, between 22% and 28% of infants were best described by Model 1, responding randomly to repeated presentations of a visual stimulus rather than showing any systematic habituation or decline over trials. Of these 25 (out of 100) infant visits, only seven failed to achieve the conventional 50%DC. In other words, nearly one fifth (18/93) of the infants who achieved this criterion were, in fact, better described by a model of random fluctuation.

Floor values $\hat{\alpha}_i$ varied across individuals as well. Approximately one third of the habituating infants (i.e., those best fit by Models 2, 3 or 4) exhibited a floor value $\hat{\alpha}_i$ at zero, while the remaining two thirds showed a non-zero value. Non-zero values of $\hat{\alpha}_i$ for habituating infants ranged from 0.71 to 19.96 seconds, with the means by model and visual stimulus as shown in Table 7.

Temporarily collapsing across the three visual stimulus conditions, Table 8 shows the number of infant visits whose responses were best characterized by each of the four models. Means, medians and standard errors of the means are also provided for $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\delta}$ within each model class.

Table 8: Mean and Median Estimated Parameter Values by Model Type

Model Number	Number of Infant Visits v	Mean Parameter Estimates (with SEM) and Median Parameter Estimates		
		$\hat{\alpha}$	$\hat{\beta}$	$\hat{\delta}$
1	25	9.85 (0.69) 7.48	--	--
2	27	--	20.66 (0.87) 19.02	0.08 (0.05) 0.05
3	40	6.69 (0.41) 5.84	43.54 (1.04) 33.06	10.45 (0.51) 1.63
4	8	9.79 (1.11) 8.51	53.74 (2.59) 54.63	15.46 (1.35) 2.04

Note: Standard errors of the mean (in parentheses) computed as in Equation 15

Substantial variation of individuals within the groups (as defined by model-based classifications) led to fewer significant mean differences between the groups. Among the three groups with non-zero estimated floor values $\hat{\alpha}_i$ (Models 1, 3 and 4), a bootstrapped oneway ANOVA failed to reject the null hypothesis of equality of means: $F = 2.059$, $p_{boot} = .1885$. Similarly, among the three groups of habituating responders (Models 2, 3 and 4) there were no significant differences in the mean $\hat{\beta}_i$, or range parameter: $F = 2.344$, $p_{boot} = .1690$. Mean comparisons of $\hat{\delta}_i$ (slope) across the three habituating groups found that infants with non-zero floor values (Models 3 and 4) decreased more quickly over trials than did infant responses best fit by Model 2: $F = 6.7792$, $p_{boot} = .0145$. These 27 infant visits (or roughly one third) whose responses were best described by a habituation model with a zero-level floor all took significantly longer to habituate than did the majority of habituating infants.

Observed differences between the mean and median parameter estimates within each of the four models reflect the non-symmetrical nature of their respective distributions. Perhaps most notable is the discrepancy between the mean and median for $\hat{\delta}_i$, particularly for Models 3 and 4. These large discrepancies are most likely artifacts of the estimation method, which used the residuals between observed and predicted values as the basis for adjusting and updating the estimated values after each iteration step. Because values of $\hat{\delta}_i$ beyond about two brought about no further change in the predicted values or

corresponding residuals for Models 2, 3 or 4 (see, for example, Equation 7), estimates for this parameter tended to explode, occasionally resulting in larger values and larger standard errors.

Effects of Stimulus Type on Habituation Behavior

Stimulus characteristics such as similarity and complexity are theorized to influence the rate at which infants habituate (e.g., Caron, 1969; Kaplan, Scheuneman, Jenkins & Hilliard, 1988), as well as affecting how stimuli are incorporated into memory (e.g., Garner, 1976; Quinn, Bhatt, Brush, Grimes & Sharpnack, 2002). Thus it was important to examine the potential impact of stimulus characteristics on model-based measurements. Because each infant saw only one class of visual stimulus, it was not possible to examine the potential effects of stimulus condition at the level of the individual infant. Group comparisons across infants who viewed different visual stimulus conditions, however, provided interesting preliminary information.

Table 9 shows the number of infants within each visual stimulus condition best described by each of the four models considered. Contingency analysis failed to reject the null hypothesis of independence between visual stimulus condition and the number of visits classified under each model: $\chi^2(6) = 5.193$ (ns).

Table 9: Model Type Classifications by Visual Stimulus Condition

Best-fit Model	Visual Stimulus Condition			
	Checkers	Optic Flow	Faces	Totals
1	7	9	9	25
2	10	11	6	27
3	13	15	12	40
4	2	1	5	8
Totals	32	36	32	100

Table 10 shows a simplified version of this same distribution of response patterns, collapsing across the different habituation models (Models 2-4):

Table 10: Habituation Response Type by Visual Stimulus Condition

Best-fit Model	Visual Stimulus Condition			
	Checkers	Optic Flow	Faces	Totals
Non-habituating (1)	7	9	9	25
Habituating (2,3,4)	25	27	23	75
Totals	32	36	32	100

Even with the reduced number of cells and their corresponding higher expected frequencies, contingency analysis still failed to reject the null hypothesis of independence between visual stimulus and response type: $\chi^2(2) = 0.333$ (ns). This suggests that the distribution, or pattern of habituation response types was independent of the visual stimulus condition.

Tables 9 and 10 suggest that the patterns of model classification were not influenced by stimulus type, at least for the three conditions examined in the current study. In other words, whether or not infants habituated was not stimulus dependent. Yet some infant behaviors did vary across the stimulus conditions, at least at the group level. Maximum estimated looking times, $\hat{w}_i = \hat{\alpha}_i + \hat{\beta}_i$, were estimated for each individual and examined at the group level. Equation 16 gives the mean maximum estimated looking time across infants, $\overline{w}$:

$$\overline{w} = \frac{\sum\limits_{\text{all } v}(\hat{\alpha}_i + \hat{\beta}_i)}{v} \tag{16}$$

where $v =$ the number of infant visits within the targeted stimulus condition. Table 11 shows these values for the three stimulus conditions, as well as the standard errors of these means, SEM($\overline{w}$), computed as in Equation 17:

$$\text{SEM}(\overline{w}) = \sqrt{\frac{\sum\limits_{\text{all } v}(\hat{\sigma}_{\hat{\beta}_i}^2 + \hat{\sigma}_{\hat{\alpha}_i}^2 + 2\hat{\rho}\hat{\sigma}_{\hat{\beta}_i}\hat{\sigma}_{\hat{\alpha}_i})}{v^2}} \tag{17}$$

where $\hat{\sigma}_{\hat{\beta}_i}$ and $\hat{\sigma}_{\hat{\alpha}_i}$ denote the bootstrap-estimated standard errors for $\hat{\alpha}_i$ and $\hat{\beta}_i$, respectively, and $\hat{\rho}$ denotes the estimated correlation between $\hat{\alpha}_i$ and $\hat{\beta}_i$.

Table 11: Maximum Estimated Looking Times $\bar{w}$ Across Stimulus Conditions

Stimulus Type	Visits v	$\bar{w}$ (in seconds)	SEM($\bar{w}$)
Non-habituating Responders (Model 1)			
Checkers	7	4.49	0.52
Optic Flow	9	7.48	0.86
Faces	9	16.39	1.66
Habituating Responders (Models 2,3,4)			
Checkers	25	20.57	1.62
Optic Flow	27	36.51	1.51
Faces	23	68.49	2.39

Note: Standard error of the mean calculated as shown in Equation 17.

These data suggest that the $\bar{w}$ were not equivalent across the three stimulus conditions, nor between habituating and non-habituating infants. With respect to stimulus condition, it would seem that infants showed different levels of maximal interest for different stimuli. Similarly one might suspect that habituating infants showed systematically more maximal interest than non-habituators, regardless of stimulus type. Both these observations were examined via bootstrapped ANOVA, leading to equivalent conclusions. Because no

significant interaction was found when examining both factors simultaneously, $F_{interaction} = 1.114$ (ns), individual oneway ANOVAs examined the two factors separately. Mean maximum estimated looking times $\bar{w}$ did vary significantly across stimulus conditions ($F = 5.61$, $p_{boot} = .0560$). Bootstrapped follow-up comparisons found that infant faces and optic flow patterns both elicited greater maximal interest than did the animated checkerboards, but that faces and optic flow patterns did not differ significantly from each other. The fact that optic flow and faces did not differ significantly was unexpected, given the observed means and standard errors shown in Table 11. Histograms of the corresponding $\hat{w}_i$ values, however, showed a single outlier in the face group that significantly elevated the observed mean, further demonstrating the potential value of the bootstrap approach, since its resampling method minimizes the potential effects of a single outlier. A bootstrapped t-test similarly found that habituating infants exhibited significantly greater maximal looking times than did non-habituating infants, regardless of stimulus type ($t = 4.998$, $p_{boot} = .0003$). These analyses suggest that the observed trends within the stimulus conditions were not confounded by the comparisons between habituating and non-habituating responders, for whom the estimated maximal interest level $\hat{w}_i$ were not precisely equivalent (i.e., $\beta = 0$ for non-habituating responders).

Habituators vs. Non-habituators: Multiple Response Types?

Table 7 provides a summary of the different observed parameter estimates for Models 1 though 4 in the various stimulus conditions, and Table 8 indicates the number of infants best described by the four habituation models altogether, collapsed across the three stimulus conditions. Evidence has been found that infants within the various stimulus conditions and model types are reacting somewhat differently. These findings suggest that discrete categories of response types may indeed exist, at least with respect to the repeating visual stimuli presented during the habituation phase. Nevertheless, it is the reactions to post-habituation stimuli that are typically of primary interest to researchers utilizing habituation techniques. Specifically, it was hypothesized that infants categorized as habituating responders would demonstrate greater dishabituation, or renewal of interest in response to a novel stimulus than would non-habituating infants. The dishabituation z-scores of Equation 13

$(z_i = \dfrac{y_i(T_i + 2) - h_i^*(T_i + 2)}{\sqrt{\hat{\sigma}^2_{h_i^*} + \hat{\sigma}^2_{h_i^*(T_i+2)} - 2\hat{\rho}(\hat{\sigma}_{h_i^*})(\hat{\sigma}_{h_i^*(T_i+2)})}})$ provided the basis for such comparisons.

A bootstrapped oneway ANOVA was used to compared each of the four types of responders (i.e., infants classified under Models 1, 2, 3 and 4) on the basis of mean dishabituation z_i-scores, collapsing across stimulus condition: $F = 2.63$ ($p_{boot} = 0.13$), which was not significant. Examining habituating responders

by themselves (Models 2, 3 and 4) similarly found no significant differences in the magnitude of the dishabituation z-scores among the three model classes: $F = 0.04$ (ns). The mean dishabituation z_i-score for each class of responders is shown in Table 12. Collapsing across the three habituation response types and comparing them to non-habituating responders, however, found that habituating infants exhibited significantly greater dishabituation, or renewal of interest to their novel stimuli (weighted mean = 2.51) than did non-habituating infants (mean = 0.28): $t = 2.51$, $p_{boot} = .006$. No significant differences in the magnitude of the dishabituation response were found between habituating infants in the three visual stimulus conditions, however: $F = 1.41$ (ns). In other words, habituating infants demonstrated roughly equivalent propensities to distinguish the novel from the familiar stimulus under each of the three visual stimulus conditions.

Table 12: Dishabituation z_i-Scores (Relative Recovery in Response to a Novel Stimulus)

Best-fit Model Type	Visits v	Mean Dishabituation z_i-score
1	25	0.28
2	27	1.98
3	40	2.91
4	8	2.23

Comparing Model-based Measures to Conventional Analyses

The fact that infants classified as habituating responders demonstrated significantly larger recovery responses to novel stimuli than did non-habituating responders lends support to the idea that these models distinguish between infants who are and are not actively engaged in the habituation process. This observation does not guarantee that model-based approaches provide any greater sensitivity to detecting such novelty responses than do conventional approaches, however. Comparisons between conventional and various types of model-based approaches can more directly examine this issue.

As previously described, conventional approaches to detecting a post-habituation novelty response use a matched pairs t-test to compare average

differences across infants in looking time responses to novel and familiar (i.e.,

habituated) stimuli presented sequentially during a post-habituation test phase

(Petrinovich & Widaman, 1984; Thomas & Gilmore, in press). This was done in

the current study as well to permit comparisons between conventional and

model-based measures of dishabituation. The post-habituation test phase in the

current study consisted of six trials alternating between the familiar and novel

stimulus, such that $y_i(T_i+1)$, $y_i(T_i+3)$, and $y_i(T_i+5)$ denoted the observed looking

time for an infant's first, third and fifth post-habituation trials (familiar

stimulus), respectively. Correspondingly, $y_i(T_i+2)$, $y_i(T_i+4)$, and $y_i(T_i+6)$

denoted the observed looking times for the second, fourth and sixth post-

habituation trials (novel), respectively. In each of the three pairs of post-

habituation trials, the familiar stimulus preceded the novel. Under the

conventional approach, matched difference scores d_i were computed for the $p = 3$

pairs of post-habituation trials such that $d_{i1}=[y_i(T_i+2)- y_i(T_i+1)]$, and similarly

for d_{i2} and d_{i3} from the second and third pairs of post-habituation trials. An

average matched difference score $\overline{d}_i = \dfrac{\sum\limits_{p=1}^{3} d_{ip}}{3}$ was also computed, and this score

served as the dependent variable for two of the comparisons to follow.

Comparison 1: The 50% Decrement Criterion and Average Matched Difference Scores

Comparison 1 used a one-tailed independent samples bootstrap t-test to examined responses from infant visits for whom the 50%DC had been achieved ($v = 93$) and those who did not achieve the criterion ($v = 7$). The dependent measure of interest was the average matched difference scores $\bar{d}_i$ computed during the post-habituation test phase. For the seven infants who did not achieve the 50%DC, the mean score was 4.33 seconds, while the mean score for infants who did achieve the criterion was 3.66 seconds: $t(98) = .154$, one-tailed $p_{boot} = .4070$ (ns). The null hypothesis of equality of means could not be rejected, meaning that the one cannot infer any difference in mean $\bar{d}_i$ scores between the two groups classified on the basis of achieving the 50%DC.

Examined separately using the conventional matched pairs t-test, both groups did show significant mean differences $\bar{d}_i > 0$. Yet what is amazing about the findings from the independent samples t-test was that the two groups were indistinguishable in terms of the magnitude of that difference. Infants who failed to achieve the 50%DC showed just as strong a decline using conventional habituation dependent measures as those who did achieve the criterion. In effect, the criterion failed to distinguish among infants on the basis of dishabituation responses, even though this is the sole purpose for which it was created.

Comparison 2: Model-based Classifications and Average Matched Difference Scores

Comparison 2 used the same dependent variable as the first comparison, the average difference score $\bar{d}_i$, computed during the post-habituation test phase on each infant visit. The independent variable in this comparison was the model-based habituation classification, collapsing across Models 2, 3 and 4 to create two groups, habituating and non-habituating responders. In this independent samples boot *t*-test, the mean $\bar{d}_i$ for the $v = 75$ habituating responders was 4.31 seconds, while for the $v = 25$ non-habituating responders, the mean $\bar{d}_i = 1.87$ seconds: $t(98) = 1.257$, $p_{boot} = .1150$, one-tailed (ns). While the mean $\bar{d}_i$ in this second comparison (2.44 seconds) was greater than in the first comparison, it would still not quite be considered significant.

Comparison 3: Model-based Classifications and Model-based Dishabituation z_i-scores

The independent variable in this comparison was the model-based habituation classification: habituating vs. non-habituating responders. Dishabituation z_i-scores, computed as in Equation 13, served as the dependent variable of focus. A bootstrapped independent samples *t*-test revealed that

habituating responders demonstrated significantly stronger dishabituation responses to the novel stimulus than did non-habituating responders: $t(98) = 2.5193$, $p_{boot} = .0097$, one-tailed.

Considered in sequence, these three comparisons show how the use of model-based classifications and model-based measures provided gradually increasing power to detect significant looking time recoveries to a post-habituation novel stimulus. In the first comparison, using conventionally-computed $\bar{d}_i$ scores and the 50%DC, no significant differences were found in the magnitude of the recovery response between infants who had or had not achieved the 50%DC, even though all infants did show recovery. Using a model-based classification to differentiate between habituating and non-habituating responders provided somewhat greater sensitivity, although the difference did not quite achieve significance using the conventional $\bar{d}_i$ as the dependent variable of focus. Only when model-based measures and classes were used to define both the independent and dependent variables was a significant difference observed. Such findings strongly suggest that researchers using habituation procedures to ask substantive questions about infant perceptual and cognitive development might benefit from the greater statistical sensitivity they provide over conventional analytical tools.

Predicting Post-Habituation Responses from Model-based Classifications

The preceding analyses suggested the existence of at least two discrete groups of infants, habituating and non-habituating, whose responses to novel visual stimuli differ. As group-based analyses, however, they do not address the issue of how well individual model-fits predict infants' responses to a novel stimulus. Specifically, one might expect that infants' looking time responses to one stimulus should resemble their responses to a similar but perceptually distinguishable stimulus presented immediately thereafter. Non-habituating infants might be expected to continue to respond randomly, whereas habituating infants would begin the process of habituating to the new stimulus. Specifically, one might expect that the shape of the fitted curve from the habituation phase, $h_i^*(t)$, might predict the shape of the curve to the novel stimulus presented during the post-habituation phase, possibly with some fatigue factor reducing the overall response levels over trials.

One way to test for this possibility was to consider the post-habituation trials (to the novel stimulus) as linear transformations of the fitted $h_i^*(t)$ responses estimated during the habituation phase for each infant, where $t = 1$, 2 and 3 are the first three habituation trials. $h_i^*(1)$, $h_i^*(2)$, and $h_i^*(3)$ can represent the fitted or expected looking time responses on the first three habituation trials, and $y_i(T_i+2)$, $y_i(T_i+4)$, $y_i(T_i+6)$ can represent the observed looking time

responses to the first, second and third presentations of the novel stimulus during

the post-habituation phase, respectively. Then a standard linear regression

model can be fit using the estimated or fitted values from the first three

habituation trials to predict the post-habituation responses to the novel stimulus.

Under this approach, the linear regression model takes the form $Y_i(t) = Bh_i^*(t) + A$,

where the parameters B and A denote the slope and intercept, respectively.

Strictly speaking, this cannot be precisely correct, since the novel stimuli

presented during the post-habituation phase were presented in alternating

sequences with the familiar stimulus, such that post-habituation trials two, four

and six were linearly regressed as functions of habituation trials one, two and

three. But as a first approximation, this seemed like a reasonable strategy.

Once standardized, the fitted slope parameter, B, might provide some

information about the rate of fatigue between the pre- and post-habituation

trials. $B \approx 1.0$ might suggest little or no systematic decrease between the two

phases, while $B \approx 0.0$ might suggest a strong decay. Fitted values of $B < 0$ would

be more theoretically difficult to interpret, indicating a negative relationship

between the pre- and post-habituation responses.

Data from the $v = 75$ habituating infant visits were individually fit in this

manner, producing separate regression models for each visit. Data from the non-

habituating infant visits could not be considered in the analysis, as there was no

variance in the fitted values of the respective $h_i^*(t)$ functions with which to explain

observed variance in the post-habituation trials. r^2 values for these linear regression models ranged from essentially 0 to 1 with a mean of .50 and a standard deviation of 0.36, as shown in Figure 11:
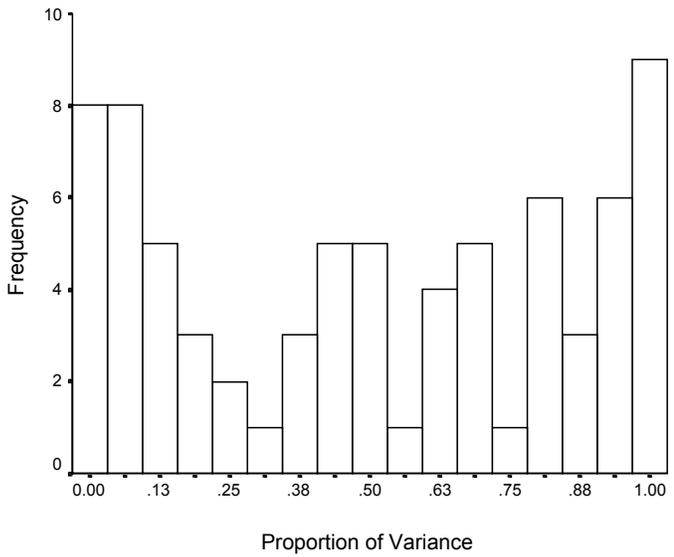
Figure 11: Proportion of Variance (r^2) Accounted for By Linear Model ($v = 75$)

Standardized B weights from the linear regression model were similarly varied, ranging from –1 to 1 with a mean of 0.30 and a standard deviation of 0.65, as shown in Figure 12:
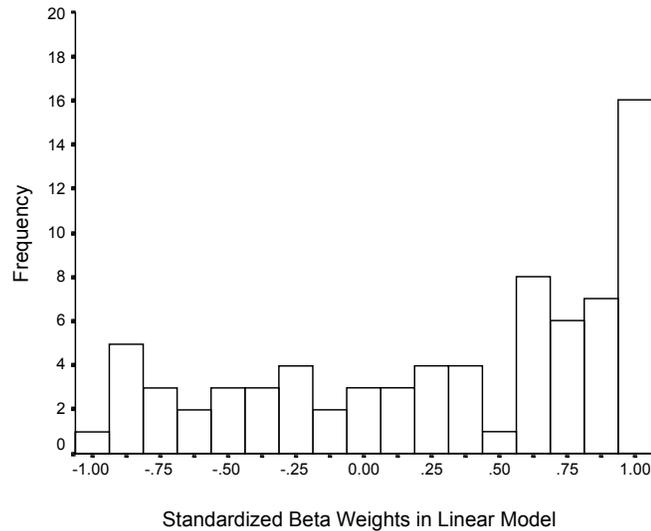
Figure 12: Beta Weights From the Linear Transformation Model ($v = 75$)

Figure 11 suggests that the linear transformation of fitted habituation responses to the observed responses to the novel stimulus was not bad as a first approximation, providing a reasonably good fit ($r^2 \geq .50$) for 45 of 75 infant visits, and a very good fit ($r^2 \geq .80$) for nearly a third. Furthermore, the distribution of standardized $\hat{B}$ weights suggested some predictive relationship between looking time responses during the habituation phase and responses to the post-habituation novel stimulus, with roughly half showing fitted $\hat{B} \geq 0.5$.

Theoretical shortcomings exist with this approach, however. Whereas during the habituation phase, the same stimulus was repeatedly presented with only a short inter-trial stimulus, the post-habituation phase consisted of six

alternating trials of the familiar and novel stimulus. This was done in order to make the current study as methodologically consistent with conventional habituation studies as possible. Nevertheless, one might posit that the intervening presentations of the familiar stimulus somehow influenced or interfered with the infant's habituation to the second (novel) stimulus.

Stability and Reliability of Model-based Classifications over Time

Previous contingency analyses failed to reject the null hypothesis of no relationship between visual stimulus type and model classification, as described earlier in this chapter in the section on "Effects of Stimulus Type on Habituation Behavior". Because no relationship or dependence could be inferred, Table 13 shows the pattern of model classifications for infants at the two visits, collapsing across the three stimulus conditions. The difference in the total numbers between visits one and two reflect the ten infants who did not return for a second visit.

Table 13: Frequency of Habituation Response Types Across Visits

Response Type	Visit Number		
	Visit One	Visit Two	Totals
Non-habituating	14	11	25
Habituating	41	34	75
Totals	55	45	100

Contingency analysis failed to reject the null hypothesis of no relationship between visit number and habituation response type, $\chi^2(1) = 0.013$ (ns), suggesting that the distribution of response types is independent of visit number. Yet this provides no information about the consistency of individual infants' responses across visits.

Tables 14 and 15 examine this issue, giving the model classifications for infants on both visits for those 45 infants who were tested twice. Table 14 shows the number of infants classified under each model on both visits:

Table 14: Model Classifications for Infants With Return Visits

Best-fit Model Visit 1	Best-fit Model Visit 2				
	1	2	3	4	Totals
1	2	3	5	0	10
2	4	3	6	1	14
3	4	3	8	2	17
4	2	1	1	0	4
Totals	12	10	20	3	45

Table 15 simplifies this presentation, collapsing across the three habituation

model types (Models 2, 3, 4).

Table 15: Habituation Response Patterns for Infants With Return Visits

Response Type Visit 1	Response Type Visit 2		
	Non-habituating	Habituating	Totals
Non-habituating.	2	8	10
Habituating.	10	25	35
Totals	12	33	45

For neither Tables 14 or 15 could the null hypothesis of classification independence be rejected: $\chi^2(9) = 3.56$ (ns) for Table 14 and $\chi^2(1) = 0.292$ (ns) for Table 15. This suggests that the issue of whether infants habituated to a visual stimulus on one occasion was independent of their habituation behavior on another occasion. Only two of forty-five infants, however, failed to demonstrate habituation behavior on at least one laboratory visit.

These analyses suggest that habituation response types, as defined by model classification, show no consistency or stability over the weeklong test interval. However, they do not provide any information about the reliability of the individual behavioral parameter estimates, such as the floor α_i, that characterize these models. Table 16 shows the Pearson's r correlations between fitted parameter estimates across visits for all 45 infants seen twice, where $\hat{\theta}_{i1}$ and $\hat{\theta}_{i2}$ refer to an infant's fitted values for a parameter on visits one and two, respectively. Considered individually, none of the behavioral parameters demonstrate much reliability across the weeklong interval except possibly for $\hat{\beta}_i$, the range parameter. The moderate correlation for $\hat{\beta}_i$ suggests that infants who showed a wide range between their maximum and minimum attention levels on one visit tended to show a greater range at the second visit. Perhaps more interesting, however, is the moderate correlation over the two visits for $\hat{\alpha}_i + \hat{\beta}_i = \hat{w}_i$, the maximum estimated looking time. This suggests that infants

who tend to look longer $(\hat{w}_i)$ on one visit tend to react similarly on the second, regardless of stimulus type shown to them.

Table 16: Pearson's r for Model Parameter Estimates Across Visits for All 45 Infants

Visit 1   Visit 2

	$\hat{\alpha}_{i2}$	$\hat{\beta}_{i2}$	$\hat{\delta}_{i2}$	$\hat{w}_{i2}$
$\hat{\alpha}_{i1}$	.071	.428	.145	.430
$\hat{\beta}_{i1}$	.104	.505	.123	.509
$\hat{\delta}_{i1}$	.066	.055	.001	-.060
$\hat{w}_{i1}$	.107	.542	-.142	.545

A similar correlational examination was done examining only the twenty-five infants who were seen twice and who habituated on both visits. This was done to determine whether the essentially random fluctuations of non-habituating infants' looking time responses might artificially deflate the observed reliability of fitted parameter estimates of habituating infants. This does not seem to have been the case, as can be seen in Table 17. The correlational patterns are roughly similar overall:

Table 17: Pearson's r for Model Parameter Estimates Across Visits for the 25 Twice-Habituating Infants

Visit 1   Visit 2

	$\hat{\alpha}_{i2}$	$\hat{\beta}_{i2}$	$\hat{\delta}_{i2}$	$\hat{w}_{i2}$
$\hat{\alpha}_{i1}$	.020	.542	-.129	.535
$\hat{\beta}_{i1}$	-.026	.622	-.238	.611
$\hat{\delta}_{i1}$	-.133	-.073	.001	.004
$\hat{w}_{i1}$	-.018	.626	-.224	.615

Perhaps most disappointing was the failure to observe any reliability of the slope parameter $\hat{\delta}_i$, which describes how quickly over trials infants reach their floor values. Rate of infant habituation, as measured by the number of trials to criterion, has previously been found to have modest predictive validity for subsequent cognitive tasks such as vocabulary later in childhood (e.g., Bornstein, 1985). Thinking this observation might reflect the particular difficulties associated with fitting the $\hat{\delta}_i$ as described in Chapter 2, a transformation was done using a cap value of 2. $\hat{\delta}_i$ were transformed into $\hat{\delta}_i^*$, where $\hat{\delta}_i^* = \hat{\delta}_i$ if $\hat{\delta}_i < 2$, and $\hat{\delta}_i^* = 2$ otherwise. Pearson's r for $\hat{\delta}_{i1}^*$ and $\hat{\delta}_{i2}^*$ was essentially the same as for the untransformed $\hat{\delta}_i$ values, however.

Pearson's r does not reflect the relative uncertainty of each parameter estimate, as denoted by standard errors. Therefore, stability $z_{\hat{\theta}_i}$ scores were computed between the fitted parameter values across visits using Equation 14

$$\left(z_{\hat{\theta}_i} = \frac{\hat{\theta}_{i1} - \hat{\theta}_{i2}}{\sqrt{\hat{\sigma}^2_{\hat{\theta}_{i1}} + \hat{\sigma}^2_{\hat{\theta}_{i2}} - 2\hat{\rho}(\hat{\sigma}_{\hat{\theta}_{i1}})(\hat{\sigma}_{\hat{\theta}_{i2}})}} \right).$$ Under this transformation, smaller $z_{\hat{\theta}_i}$ scores

indicated smaller deviations or greater reliability for each pair of fitted values.

The observed distributions of stability $z_{\hat{\theta}_i}$ scores were then evaluated under the

Kolmogorov-Smirnov z-test to determine whether the observed responses were

consistent with a standard normal distribution.

Figure 13 shows the sample distribution of the stability $z_{\hat{\theta}_i}$ scores between

times one and two for the $\hat{\alpha}_i$ estimates of the infants whose model classifications

were consistent across visits (habituating both times or not habituating both

times):
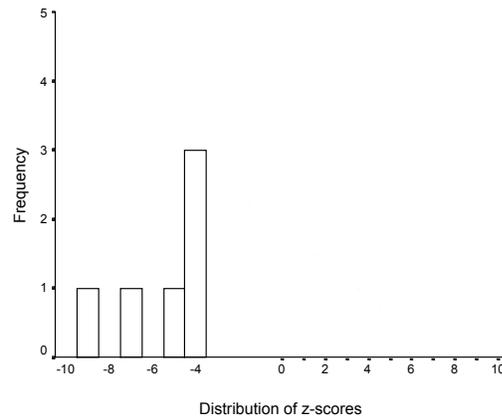
Figure 13: Distribution of the 22 $z_{\hat{\theta}_i}$ Scores for $\hat{\alpha}_i$ With Mean = -0.99, Standard Deviation = 3.69

The one-sample Kolmogorov-Smirnov z-test rejected the null hypothesis of standard normality for the observed sample distribution: $z_{ks} = .4545, (p = .0001)$.

Figure 14 shows the corresponding distribution of the stability $z_{\hat{\theta}_i}$ scores for the fitted range parameter $\hat{\beta}_i$, though the single extreme outlier at $z = $ -53 impacts the observed sample mean and standard deviation:
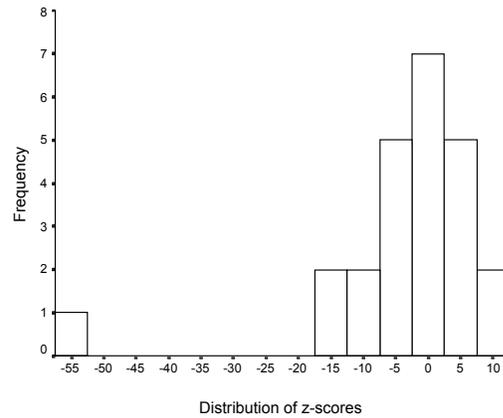
Figure 14: Distribution of the 24 $z_{\hat{\theta}_i}$ Scores for $\hat{\beta}_i$ With Mean = -3.89, Standard Deviation = 12.26

Figure 15 shows the distribution of stability $z_{\hat{\theta}_i}$ scores for $\hat{\beta}_i$ after removal of the outlier:
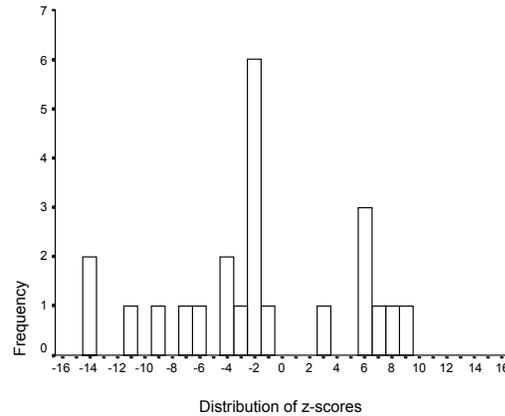
Figure 15: Distribution of the 23 $z_{\hat{\theta}_i}$ Scores for $\hat{\beta}_i$ With Mean $= -1.77$, Standard Deviation $= 6.66$

Similarly, the one-sample Kolmogorov-Smirnov test rejected the null hypothesis of standard normality for the observed sample distribution: $z_{ks} = .6247$ ($p < .0001$).

Figure 16 provides the corresponding sample distribution of stability $z_{\hat{\theta}_i}$ scores for the observed $\hat{\delta}_i$ parameters over times one and two. For this distribution, however, the Kolmogorov-Smirnov just barely failed to reject the null hypothesis of standard normality for the observed distribution of $z_{\hat{\theta}_i}$ scores: $z_{ks} = .2462$ ($p = .0919$).
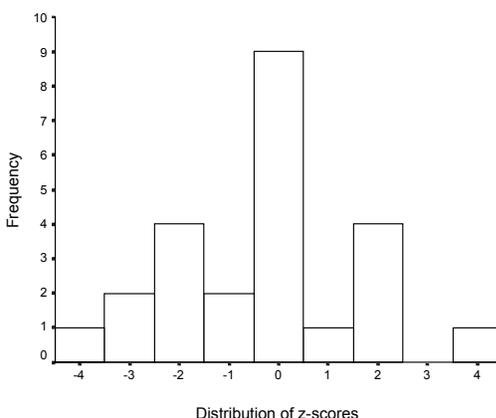
Figure 16: Distribution of the 24 $z_{\hat{\theta}_i}$ Scores for $\hat{\delta}_i$ With Mean = -0.33, Standard Deviation = 1.92

The observed standard deviations in Figures 13 and 15, as well as the corresponding z_{ks} scores, would suggest that the variability for $\hat{\alpha}_i$ and $\hat{\beta}_i$ across visits was much greater than one might predict under the standard normal distribution. This was nearly the case for $\hat{\delta}_i$ as well, although statistical significance at the conventional alpha = .05 level was not quite met. The rough symmetry shown in Figures 13, 15 and 16 do suggest the possibility of a normal distribution, however, even if the standard normal was rejected. Evidence for non-standard normality was further supported when the Kolmogorov-Smirnov was repeated, using test parameters generated from the observed sample moments. Normality was not rejected for any of the three distributions.

All three $z_{\hat{\theta}_i}$ score distributions were roughly normal and with similar
sample means and variances. The non-parametric Friedman test examining the
relative ranks of the three samples failed to reject the null hypothesis that the
average case rankings within the three samples were the same: $\chi^2(2) = 1.3$ (ns).
Similarly, bootstrap hypothesis tests failed to reject the null hypothesis of
equality of means for the three distributions: $F = 1.50$ (ns). Taken together, one
may conclude that none of these three parameters, on average, demonstrated any
greater reliability than the others over the weeklong interval. Independence of
habituation responses across visits and the relative lack of reliability for
individual parameter estimates suggest little stability or reliability across the
visits, showing large differences both within and across individuals on the
observed habituation behaviors.

2-, 3- and. 4-Parameter Habituation Models

Dishabituation z_i-scores ($z_i = \dfrac{y_i(T_i+2) - h_i^*(T_i+2)}{\sqrt{\hat{\sigma}_{h_i^*}^2 + \hat{\sigma}_{h_i^*(T_i+2)}^2 - 2\hat{\rho}(\hat{\sigma}_{h_i^*})(\hat{\sigma}_{h_i^*(T_i+2)})}}$) for the three
classes of habituating infants (Models 2, 3, and 4) were all roughly equivalent
and were all significantly greater than for non-habituating infants. Given such
equivalence, at least with respect to detecting a post-habituation response, one
might question whether fewer models might equally suffice. Particularly of

interest was the issue of whether the 4-parameter model was necessary to differentiate between habituating and non-habituating responses.

While Model 1 predicted no change for infants' looking time responses over trials, all three habituating models did. In order to determine whether all habituating models accounted equally well for any observed changes in infants' responses, a proportion of variance score ε_i was computed for each infant using the sample variance of the observed responses $y_i(t)$ and the residual error under the best-fit model $h_i^*(t)$: $\varepsilon_i = 1 - \dfrac{\hat{\sigma}_{h_i^*}^2}{\hat{\sigma}_{y_i}^2}$. $\hat{\sigma}_{y_i}^2$ was computed as the conventional observed sample variance for each infant i during the habituation phase:

$$\hat{\sigma}_{y_i}^2 = \left[\frac{\sum_{t=1}^{T_i}(y_i(t) - \overline{y}_i)^2}{T_i - 1} \right]^2 .$$ $\overline{y}_i$ was the observed sample mean for each infant i:

$$\overline{y}_i = \frac{\sum_{t=1}^{T_i} y_i(t)}{T_i}.$$ Error variance under the best-fit model $\hat{\sigma}_{h_i^*}^2$ was computed using the estimated model residual error: $\hat{\sigma}_{h_i^*}^2 = \left[\dfrac{\sum_{t=1}^{T_i}(y_i(t) - h_i^*(t))^2}{T_i - k} \right]^2 .$ T_i denoted the number of habituation trials for infant i, and k gave the number of estimated parameters in the best-fit model. The quantity $\dfrac{\hat{\sigma}_{h_i^*}^2}{\hat{\sigma}_{y_i}^2}$ estimated the proportion of observed sample variance not predicted by the best-fit model, while $\varepsilon_i = 1 - \dfrac{\hat{\sigma}_{h_i^*}^2}{\hat{\sigma}_{y_i}^2}$ gave the remaining variance, or the proportion that was accounted for under the model.

Although the estimated models were nonlinear, in no case within the current

study did ε_i give rise to negative values (i.e., $\dfrac{\hat{\sigma}^2_{h_i^*}}{\hat{\sigma}^2_{y_i}}$ was always less than one).

Table 18 shows the ordinary sample means and standard errors of the

mean for the proportions of variance $\bar{\varepsilon}$ accounted for by each best-fit model type

except for Model 1, which predicted no change over trials. In fact, the standard

errors reported in Table 18 cannot be strictly correct, since it would require the

calculation of the bivariate Taylor series, taking into account the variance and

potential covariance associated with $h_i^*(t)$ and $y_i(t)$. But the ordinary sample

SEM suffices for illustrative purposes.

Table 18: Mean Proportions of Variance $\bar{\varepsilon}$ Accounted for by Model

Model Number	Visits v	$\bar{\varepsilon}$	SEM($\bar{\varepsilon}$)
1	25	n/a	---
2	27	.58	.04
3	40	.88	.02
4	8	.68	.08

A bootstrapped oneway analysis of variance was used to estimate the

sampling distribution of $\bar{\varepsilon}$, revealing a significant difference ($F = 24.8587$, $p_{boot} <$

.0001) in how well each model accounted for observed variance in infants'

responses. Bootstrapped post-hoc mean comparisons showed that when Model 3 was selected as the best-fit model, it accounted for significantly more observed variance in looking time responses than did either Models 2 or 4, which did not differ from each other.

Given this apparent advantage of Model 3 over Models 2 and 4, infants initially classified under Models 2 and 4 were re-evaluated to see how many of them would still be considered habituating responders under a two model system that only included Models 1 and 3. Under this narrower classification scheme, the original sample of $v = 100$ infant visits would be re-classified as shown in Table 19, with roughly 60% of the infant visits classified as habituating and 40% classified as non-habituating. Of the eight infants initially classified under Model 4, seven were re-classified as non-habituating, while eleven of the Model 2 infants were re-classified as non-habituating.

Table 19: Re-Classification of $v = 100$ Infant Visits Under a Two Model Schema

Original Classification	New Classification		
	Model 1	Model 3	Total
Model 1	25	0	25
Model 2	11	16	27
Model 3	0	40	40
Model 4	7	1	8
Total	43	57	100

Table 13 may be re-written to reflect the narrower classification scheme for the $n = 45$ infants who were twice tested. This is shown in Table 20. The overall classification pattern under the narrower scheme is comparable to the original, with approximately 56-58% of all infants exhibiting habituation responses on any given day, and roughly 75% of all infants habituating at least once. As before, the null hypothesis of independence of classification across the two visits could not be rejected: $\chi^2(2) = 2.409$, $p = .121$.

Table 20: Response Types Across Visits Under a Two Model Schema for $n = 45$ Infants

Response Type Visit 1	Response Type Visit 2		
	Non-habituating	Habituating	Totals
Non-habituating	11	8	19
Habituating	9	17	26
Totals	20	25	45

Re-examining novelty dishabituation z_i-scores on the basis of this new classification actually yielded a larger difference between habituating and non-

habituating responders, however: bootstrapped $t = 2.827$ (vs. $t = 2.513$ under the initial classification). This observed change is primarily attributable to the infants initially classified under Model 4. Figure 17 shows the distribution of dishabituation z_i-scores for these infants:
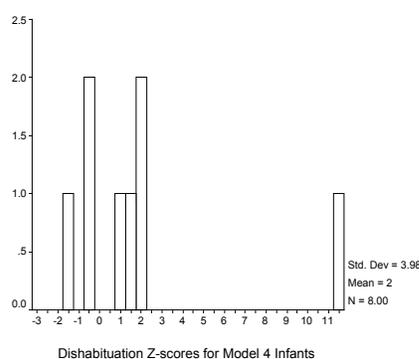
Figure 17: Histogram of Dishabituation z_i-Scores for Model 4 Visits

The same seven infants who were re-classified under Model 1 show relatively weak dishabituation responses ($|z_i| \leq 2$), while the single infant re-classified under Model 3 showed a particularly strong dishabituation response. These observations suggest that the 4-parameter model is unnecessary for distinguishing between habituating and non-habituating responses, and may actually constitute "over-fitting of the data"

Chapter 4

Discussion

Visual habituation has grown increasingly popular over the past four decades as a behavioral tool for answering questions about perceptual and cognitive development during infancy. Yet during that time, little effort has been devoted towards improving the criteria by which infant habituation is measured. The most common habituation criterion currently in use, the fifty percent decrement criterion, has remained essentially unchanged since its inception in the early 1970's. Yet shortcomings exist with this criterion – some previously acknowledged and some not – that may be overcome with the use of model-based measures of habituation. The 50%DC was first proposed as a means of capturing and equating infants exhibiting different rates and degrees of habituation to a repeating stimulus. The current study, however, demonstrated how model-based measures and classifications systems outperformed the 50%DC, both in distinguishing habituating and non-habituating infants, and in detecting significant post-habituation novelty responses. And while the mathematical complexity of model-based measures may have rendered them impractical in the

sixties and early seventies when habituation methods were developed, current availability of fast, inexpensive computer technology renders such issues moot.

Infants in the current study were found to exhibit large behavioral differences, some showing large decreases in looking time over repeated trials, others showing smaller changes, and some showing no systematic decreases over trials. Among infants classified as habituating, roughly one third were described as having a zero floor estimate ($\hat{\alpha}_i = 0$), while about two thirds showed $\hat{\alpha}_i > 0$, ranging from 0.71 to 19.96 seconds.

Patterns of habituation response were essentially independent across visits separated by a weeklong interval, although essentially all (43/45) infants habituated at least once. Independence across visits as well as poor reliability of model parameter estimates suggested that habituation behaviors on any given day varied greatly, and may be more strongly influenced by immediate or "state" arousal levels than by longer term temperamental factors.

Habituating infants (i.e., best fit by one of Models 2, 3 or 4), tended to show stronger dishabituation or recovery responses to a novel stimulus than did non-habituating (Model 1) responders. No significant differences were found among infants classified under Models 2-4. Comparisons between conventional analytical approaches and model-based measures showed that using model-based classifications and measures provided increased sensitivity to detect a post-habituation recovery response to a novel stimulus.

Comparisons across the three types of visual stimuli indicated that infant faces and optic flow patterns elicited greater maximal interest among participating infants than did the animated checkerboard patterns, but that patterns of habituation response were independent of stimulus type. Furthermore, when measured by dishabituation z_i-scores, recovery of interest did not vary across the three stimulus conditions.

Alternatives to the 50% Decrement Criterion

Most researchers employ habituation techniques as tools for studying substantive questions about infant cognitive and perceptual development. In such cases, infant control procedures are generally seen as useful methods for achieving comparable levels of habituation amongst individuals before exposing them to a post-habituation stimulus (Lavoie & Desrochers, 2002). It is the response to this post-habituation stimulus that is usually the real focus of interest to researchers. Yet the 50%DC, the current "gold standard" by which habituation is measured under infant control procedures, cannot distinguish between infants who have truly habituated and those who merely achieve the criterion by chance. In the current study, for example, responses from 18 infant visits out of a total of 100 achieved the 50%DC when their looking times were in fact better characterized by non-systematic fluctuation about a mean.

Not only does the 50%DC fail to distinguish between habituating and non-habituating infants, it fails to elicit or measure "equivalent" levels of habituation, even among infants whose responses are systematically decreasing over trials. The very idea of equating implies that each infant can be brought to roughly the same proportional decrement, relative to individual maximum and minimum levels. This can only be done in one of two ways: either all infants must be assumed to have the same scale (i.e., maximum and/or minimum response levels), or scales must be individually estimated for each infant. Under the 50%DC, a maximum, or baseline response is estimated individually for each infant based on the first few looking time responses. Yet a common floor – at zero – must also be assumed for all infants in order for the criterion to be a reasonable measure. The current study, however, found that nearly two thirds of all habituating infants were better characterized by non-zero floors ($\hat{\alpha}_i > 0$) when this parameter was individually estimated through model-based methods.

Thomas and Gilmore (in press) propose an alternative to the 50%DC derived from Model 3 in the current study, but this too is flawed. Under their criteria, model parameters are estimated online (i.e., during infant testing) and used to compute the appropriate termination trial for any desired proportion of habituation, p. Specifically, if m_i and f_i denote an infant i's maximal and floor level looking time responses, respectively, and $h_i(t_{pi})$ denotes the habituation function giving infant i's looking time on trial t corresponding to p, then

$$p_i = \frac{m_i - h_i(t_{pi})}{m_i - f_i}.$$ And since under Model 3, the maximum $m_i = (\alpha_i + \beta_i)$ happens

on trial t $=1$, while $\lim_{t\to\infty} h_i(t) = \alpha_i = f_i$, then after substitution, they present

their Equation 6, re-expressed here as Equation 18:

$$p_i = 1 - \exp[-\delta_i(t_{pi} - 1)^2] \tag{18}$$

Replacing p_i with p in their Equation 6 and solving for t gives their Equation 7,

re-expressed here as Equation 19:

$$t_i(p, \delta_i) = \sqrt{\frac{\log(\frac{1}{1-p})}{\delta_i}} + 1 . \tag{19}$$

Equation 19 can be computed for real data using the corresponding parameter

estimates. For example, an infant i with an estimated delta parameter $\hat{\delta}_i = 1.5$

would habituate to a relative proportion $p = 50\%$ on trial $\hat{t}_i(p, \hat{\delta}_i) = \sqrt{\frac{\log\frac{1}{1-p}}{\hat{\delta}_i}} + 1$

$= \sqrt{\frac{\log\frac{1}{1-.5}}{1.5}} + 1 = 1.68$. This value, rounded to the next higher integer, predicts

that 50% relative habituation for this infant would occur at approximately trial

2. Pilot trials for the current study demonstrated that online estimation of the

model parameters necessary for computing this criterion could be done, after a

minimum of six habituation trials, with only a minimal delay (1-2 seconds)

between trials when only two models (i.e., Models 1 and 3) were evaluated.

Practical flaws limit the viability of Equation 19 as an online stopping criterion, however. In the current study, for example, $\hat{\delta}_i$ estimates for the $v = 75$ habituating infant visits ranged from 0.01 to 41.87 with a median $\hat{\delta}_i = 0.986$. Using the median estimated value, roughly 50% of all habituating infants would achieve 50% relative habituation by trial $t = 1.84 \approx 2$. Yet it is impossible to generate estimates under the 3-parameter model using only two observations. One might consider proportions $p >> .50$, though it is worth noting that for 100% relative habituation ($p = 1$), Equation 19 is undefined. This is a consequence of the conditions under the model (i.e., "soft landings") such that there exists no trial t under the model for which the expected value at $t = \alpha_i$. To be sure, one can get arbitrarily close to 100% relative habituation ($p = .9$, for example). But even with $p = .9$ under Equation 19, the median $t_i(p, \hat{\delta}_i) = 2.52 \approx 3$ in the current study (based on a median $\hat{\delta}_i = 0.986$), which is still too few trials for generating reliable estimates under the three-parameter model.

Rather than seeking a model-based stopping criterion to stand in lieu of the 50%DC, a simpler alternative would be to return to the original meaning of the term "infant control" (Horowitz, et al., 1972b), in which a fixed number of habituation trials of varying duration are used for all infants. This alternative would require identification of some optimal *fixed* number of trials that would permit accurate model parameter estimates, given the observed variation in those parameters across and within infants. Although this might seem counter-

intuitive, a fixed number of trials would eliminate the need for online computation of any habituation criterion, greatly simplifying the testing procedure. If it could be shown that a fixed number of trials were sufficient to categorize infants as habituating or non-habituating, as well as to bring habituating infants to nearly complete habituation, relative criteria such as the 50%DC would be unnecessary.

An optimal number of trials, if it existed, would have to be based on several considerations. First of all, such a number would need to provide sufficiently many observations to allow estimation of all model parameters. In the current setting, $k+1$ observations were required to generate BIC estimates for any model, where k gives the number of fitted parameters in the model. So in the current study, at least five observations would have been necessary.

Beyond identifiability considerations, however, an optimal fixed number of trials would need to allow infants with varying rates of habituation δ_i to reach some relative proportion of habituation p arbitrarily close to 1. Given p and $\hat{\delta}_i$, this number can be individually computed under Equation 19. Within the current study, Figure 18 examines this question for the distribution of $v = 75$ habituating infant visits under a relative proportion of habituation $p = .9$:
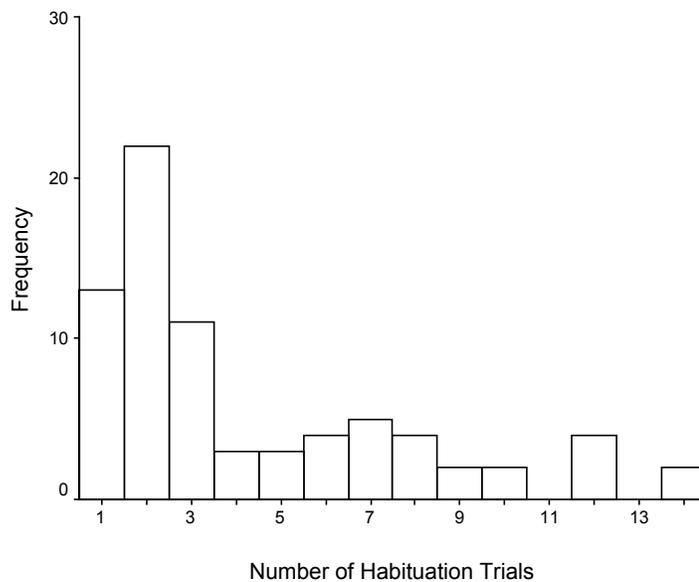
Figure 18: Estimated Number of Trials to Reach 90% Relative Habituation Under Equation 19

The vast majority of habituating responders ($62/75 \approx 80\%$) achieved $p \geq .9$ relative habituation by trial $t = 7$, and almost all of them would be expected to achieve this level of relative habituation by trial $t = 10$.

Figure 18 suggests that nearly all habituating infants in the current study would have been expected to achieve a relative proportion of habituation arbitrarily close to 1 ($p = .9$) within a fairly small number of fixed trials. Still, the figure provides no information about how many habituation trials were necessary to stably classify infants as "belonging" under Models 1-4. In the current study, all infants saw between six and fifteen habituation trials,

contingent upon achieving the 50%DC. Nevertheless, in many cases infants could have been classified using the model-based approach with fewer trials. It was previously observed that identifiability issues within the current model framework required no fewer than five habituation trials per infant to simultaneously compare the four models under consideration. An infant classified under Model 1 (non-habituating), however, could theoretically be fit with as few as two observations. Infants classified under Models 2 and 3 could similarly have been fit with as few as three or four trials. What is not known, however, is whether each infant's model-based classification using fewer than the maximum number of available habituation trials would result in the same classification as was achieved using all available habituation trials. The number of habituation trials necessary to achieve this "stable" model classification would be of use in recommending an appropriate number of fixed trials for future studies.

In order to examine this issue, each infant's sequence of looking time responses during habituation were sequentially re-fit under each model, one observation at a time. For example, Model 1 was fit to each infant's first two trials and a corresponding estimate of BIC was generated. The third trial was then included and the data were re-fit to Models 1 and 2, saving both BIC estimates. The fourth trial was added and the data were re-fit to Models 1, 2 and 3, saving the corresponding BIC estimates for each. In this manner, it was possible to observe how each infant would have been classified using fewer than

the total number of available trials, and to determine how many trials were

necessary for each infant to achieve his/her "final" classification. Figure 19

shows this distribution, indicating that roughly 90% of all infant visits in the

current study achieved a stable model-based classification within seven trials.
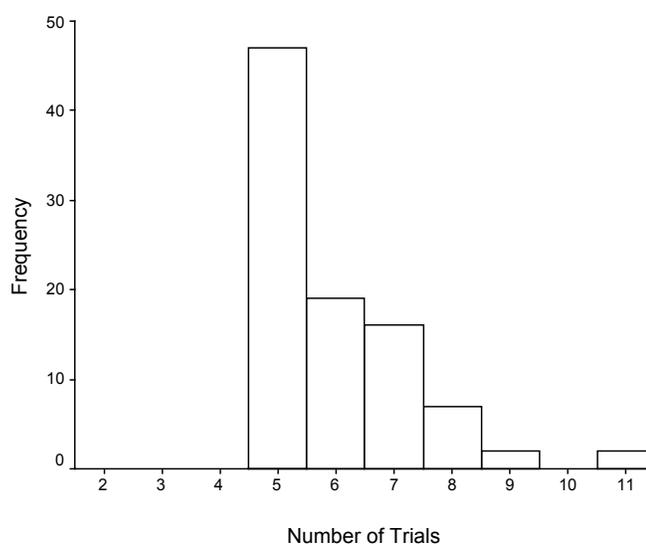
Figure 19: Number of Habituation Trials Necessary to Achieve "Final" Classification

Considered simultaneously, Figures 18 and 19 suggest that the vast

majority of infants within the current study would have been stably classified as

habituating or non-habituating within seven trials, and that most habituating

infants would be very near complete relative habituation within those same seven

trials. Consequently, one might conclude that a fixed number of seven

habituation trials would be sufficient for the vast majority of infants, at least within the current study.

Shortcomings Associated with Model-Based Measures

Increased power to detect post-habituation novelty responses, coupled with potentially simplified testing procedures through the elimination of online habituation criteria without loss of the ability to detect and measure individual differences constitute compelling advantages of model-based measures over conventional infant control procedures utilizing the 50%DC. Nevertheless, there are certain difficulties associated with these approaches.

Estimation Procedures

The models and most of the statistical analyses described within the current study were written using MATLAB scripting language rather than through pre-built "point and click" functions common to many standard statistical programs. Some packages (e.g., SPSS, SAS, Matlab) do provide functionality for non-linear regression, as well as bootstrap procedures for estimating standard errors, however the models and the BIC criterion used for selection must be custom-built. Included within the appendices of this

manuscript are the critical algorithms used for estimation and statistical inference in the current study, written in Matlab script. Generally speaking, this script is similar to S+ and C++, and would be relatively adaptable. Eventually it is hoped to provide an executable program that performs all estimation procedures, requiring no programming ability by the user. Meanwhile, however, the current Matlab scripts will be made freely available upon request.

Reliability/Stability Issues

The relative lack of stability/reliability of model-based measures across the weeklong interval within the current study are somewhat at odds with prior research demonstrating moderate reliability of habituation behaviors (e.g., Colombo, 1993; Bell, Slater, ALSPAC Study Team, 2002; Lavoie & Desrochers, 2002). Contingency analyses based on Tables 14 and 15 (illustrating the patterns of classification for infants tested twice) fail to reject a null hypothesis of independence across visits, and the correlations across visits of individual parameter estimates (e.g., Table 17) show only moderate associations at best. The large standard errors associated with various model parameter estimates in the current study, as can be seen in Table 7 (e.g., $\hat{\delta}_i$), limit the utility of many reliability measures such as Pearson's r. But even when standard errors are taken into account through the use of the stability $z_{\hat{\theta}_i}$ scores (Equation 14) as

shown in Figures 13, 15, and 16 for $\hat{\alpha}_i$, $\hat{\beta}_i$, and $\hat{\delta}_i$, respectively, variability was still greater than one might hope. Potential efforts to improve the precision of model-based measures include combining information from multiple visits (for example, a linear combination across visits, inversely weighted by their respective standard errors) to generate a single point estimate. Empirical Bayes shrinking methods (e.g., Casella, 1985; Thomas, 1993) also provide a means for generating improved measures by drawing upon information from the other individuals in the sample.

Even though there was little stability or reliability over the weeklong interval, the fact that model-based classifications significantly predicted post-habituation novelty responses on the same day, suggests that these models are indeed tapping into the underlying perceptual and/or memory processes involved with habituation, but that such processes may not remain consistent across visits. Alternatively, it may be that other transient "state" factors (e.g., mood, arousal levels, physical comfort levels) may interfere with the habituation process, and that these conditions are not consistent across time.

Further Directions

Potential areas for future work might include efforts to improve the precision of model-based habituation measures through the incorporation of

parent report information related to infants' arousal states. If, in fact, transient internal states such as irritability or fatigue are predictors of habituation performance on any given day, then asking parents about their infants' recent sleep or feeding patterns might provide useful information about which infants would be most likely to habituate to a stimulus. Furthermore, accounting for such variability might help to provide more reliable habituation measures across visits. Other potential covariates include EEG (Kaufman, Csibra & Johnson, 2003) and heart rate (Hansen, Johnsen & Thayer, 2003) measures, both thought to correspond to the formation of internal mental representations, memory and attention.

Other directions include the development of a joint model of the habituation and post-habituation looking time responses. Such efforts would further examine the potential relationships between infant habituation to the familiar and novel stimuli presented on a single experimental visit. In the current study, the novel stimulus was shown during the post-habituation phase in an alternating sequence in conjunction with the previous "familiar" stimulus from the habituation phase. This was done in order to provide maximum comparability to conventional infant control procedures, which use alternating sequences of familiar/novel trials during the post-habituation phase as the basis for matched pairs t-tests. Using an alternating sequence, however, likely

interferes with the habituation of the second stimulus, making any direct pre- and post-habituation comparisons between fitted models less appropriate.

Building on the current results, a new study could be done using the original infant control paradigm (i.e., fixed number of habituation trials of variable duration) to examine the relationship between model fits obtained to the habituation stimulus and to the post-habituation stimulus. In such a study, infants would view a habituating stimulus for a fixed number of trials, then view a similar but perceptually distinctive novel stimulus for the same number of fixed trials. Each sequence of looking time responses would be separately fit and classified under Models 1 and 3, and relationships between the fitted parameter estimates between the two phases could be sought. Specifically, one might predict that pre- and post-habituation responses would follow the same general model type, with possibly an attenuation or "fatigue" factor applied to the post-habituation responses to reflect infants' limited endurances. If functional relationships between pre- and post-habituation behavior were identified, it might be possible to predict individual post-habituation looking time responses with even greater accuracy than was possible in the current study, rendering habituation techniques even more useful to researchers examining substantive questions about perceptual and cognitive development.

Bibliography

Arterberry, M.E., & Yonas, A. (2000). Perception of three-dimensional shape specified by optic flow by 8-week old infants. Perception and Psychophysics, 62(3), 550-556.

Ashmead, D.H., & Davis, D.L. (1996). Measuring habituation in infants: An approach using regression analysis. Child Development, 67, 2677-2690.

Baillargeon, R. (1991). Reasoning about the height and location of a hidden object in 4.5- and 6.5-month old infants. Cognition, 38, 13-42.

Bell, J.C., Slater, A., & ALSPAC Study Team (2002). The short-term and longer-term stability of non-completion in an infant habituation task. Infant Behavior and Development, 25, 147-160.

Bogartz, R.S. (1965). The criterion method: Some analyses and remarks. Psychological Bulletin, 64, 1-14.

Bornstein, M.H. (1985a). Habituation of attention as a measure of visual information processing in human infants: Summary, systematization, and synthesis. In G. Gottlieb & N.A. Krasnegor (Eds.), Measurement of Audition and Vision in the First Year of Post-natal Life: A Methodological Overview (pp. 253-300). Norwood, NJ: Ablex Publishing Corporation.

Bornstein, M.H. (1989). Stability in early mental development: From attention and information processing in infancy to language and cognition in childhood. In M.H. Bornstein, & N.A. Krasnegor (Eds.), Stability and Continuity in Mental Development (pp. 147-170). Hillsdale, NJ: Erlbaum.

Bornstein, M.H. (1998). Stability in mental development from early life: Methods, measures, models, meanings and myths. In F. Simeon & G.E. Butterworth (Eds.), The Development of Sensory, Motor and Cognitive Capacities in Early Infancy: From Sensation to Cognition (pp. 299-331). Hove, United Kingdom: Psychology Press.

Bornstein, M.H., Ferndinandsen, K., & Gross, C.G. (1981). Perception of symmetry in infancy. Developmental Psychology, 17(1), 82-86.

Bornstein, M.H., & Tamis-LeMonda, C.S. (1994). Antecedents of information-processing skills in infants: habituation, novelty responsiveness, and cross-modal transfer. Infant Behavior and Development, 17, 371-380.

Boyd, E. F. (1974). Visual fixation and voice discrimination in 2-month old infants. In F.D. Horowitz (Ed.), Visual Attention, Auditory Stimulation, and Language Discrimination in Young Infants. Monographs of the Society for Research in Child Development, 39, (5-6, Serial No. 158), pp. 63-77.

Brainard, D.H. (1997). The psychophysics toolbox. Spatial Vision, 10, 443-446.

Branch, M.A., Coleman, T.F., & Li., Y (1999). A subspace, interior, and conjugate gradient method for large-scale bound-constrained minimization problems. SIAM Journal on Scientific Computing, 21(1), 1-23.

Caron, R. (1969). Degree of stimulus complexity and habituation of visual fixation in infants. Psychonomic Sciences, 14(2), 78-79.

Casella, G. (1985). An introduction to empirical Bayes data analysis. American Statistician, 39, 83-87.

Cohen, L.B. (1976). Habituation of visual attention. In T.J. Tighe & R.N.

Leaton (Eds.), <u>Habituation: Perspectives from Child Development,</u>

<u>Animal Behavior, and Neurophysiology</u> (pp. 207-238). Hillsdale, NJ:

Erlbaum.

Cohen, L.B., & Menten, T. (1981). The rise and fall of infant habituation.

<u>Infant Behavior and Development, 4</u>, 269-280.

Colombo, J. (1993). <u>Infant Cognition: Predicting Later Intellectual Functioning</u>.

Newbury Park, CA: Sage Publications.

Colombo, J., & Mitchell, D.W. (1990). Individual differences in early visual

attention: Fixation time and information processing. In J. Columbo, & J.

Fagen (Eds.), <u>Individual Differences in Infancy: Reliability, Stability,</u>

<u>Prediction</u> (pp. 193-227). Hillsdale, NJ: Erlbaum.

Dannemiller, J.L. (1984). Infant habituation criteria: I. a monte carlo study of

the 50% decrement criterion. <u>Infant Behavior and Development, 7</u>, 147-

166.

Dodd, C., & Lewis, M. (1969). The magnitude of the orienting response in

children as a function of changes in color and contour. <u>Journal of</u>

<u>Experimental Child Psychology, 8</u>, 296-305.

Efron, B., & Tibshirani, R.J. (1993). An Introduction to the Bootstrap. New

    York: Chapman and Hall.

Everitt, B.S. (1987). Introduction to Optimization Methods and Their

    Applications in Statistics (pp.23-24). London: Chapman and Hall.

Fagan, III, J.F. (1984). The relationship of novelty preferences during infancy to

    later intelligence and later recognition memory. Intelligence, 8, 339-346.

Fantz, R.L. (1964). Visual experience in infants: Decreased attention to familiar

    patterns relative to novel ones. Science, 146, 668-670.

Feldman, R., & Mayes, L.C. (1999). The cyclic organization of attention during

    habituation is related to infants' information processing. Infant Behavior

    and Development, 22(1), 37-49.

Freedle, R. (1971). A stimulus similarity scale for temporal measures of

    attention in infants and children. Developmental Psychology, 4, 240-247.

Garner, W.R. (1976). Interaction of stimulus dimensions in concept and choice

    processes. Cognitive Psychology, 8, 98-123.

Gilmore, R.O., & Rettke, H. (2003). Four month olds' discrimination of optic

    flow patterns depicting different directions of observer motion. Infancy,

    4(2), 177-200.

Gilmore, R.O., Baker, T.J., & Grobman, K.H. (in press). Stability in young
    infants' perception of optic flow. <u>Developmental Psychology</u>.

Gilmore, R.O. & Thomas, H. (2002). Examining individual differences in infants'
    habituation patterns using objective quantitative techniques. <u>Infant
    Behavior and Development, 153,</u> 1-14.

Hansen, A.L., Johnsen, B.H., & Thayer, J.F. (2003). Vagal influence on working
    memory and attention. <u>International Journal of Psychophysiology, 48(3),</u>
    263-274.

Hogg, R.V., & Tanis, E.A. (1997). <u>Probability and Statistical Inference.</u> (pp.
    609-614). Upper Saddle River, NJ: Prentice Publishing.

Horowitz, F.D., Paden, L., Bhana, K., & Self, P. (1972b). An infant-control
    procedure for studying infant visual fixations. <u>Developmental Psychology,
    7(1)</u>, 90.

Johnson, S.P., & Aslin, R.N. (2000). Infants' perception of transparency.
    <u>Developmental Psychology, 36(6)</u>, 808-816.

Johnson, S.P., Bremner, J.G., Slater, A.M., Mason, U.C., & Foster, K. (2002).
    Young infants' perception of unity and form in occlusion displays. <u>Journal
    of Experimental Child Psychology, 81</u>, 358-374.

Kagan, J., Linn, S., Mount, R., Reznick, J.S., & Hiatt, S. (1979). Asymmetry of inference in the dishabituation paradigm. Canadian Journal of Psychology, 33(4), 288-308.

Kaplan, P.S., Scheuneman, D., Jenkins, L., & Hilliard, S. (1988). Sensitization of infant visual attention: Role of pattern contrast. Infant Behavior and Development, 11, 265-276.

Kaplan, P.S., & Werner, J.S. (1986). Habituation, response to novelty, and dishabituation in human infants: Tests of a dual-process theory of visual attention. Journal of Experimental Child Psychology, 42, 199-217.

Kaplan, P.S., & Werner, J.S. (1987). Sensitization and dishabituation of infant visual attention. Infant Behavior and Development, 10, 183-197.

Kaplan, P.S. (1990). Habituation, sensitivity and infant visual attention. Advances in Infancy Research, 6, 61-100.

Kaufman, J., Csibra, G., & Johnson, M.H. (2003). Representing occluded objects in the human infant brain. [Electronic Version] Proceedings of the Royal Society of London (Supplement: Biology Letters), S1-S4.

Kirkham, N.Z., Slemmer, J.A., & Johnson, S.P. (2002). Visual statistical learning in infancy: evidence for a domain general learning mechanism. Cognition, 83, B35-B42.

Laub, K.W., & Bhana, K. (1974). Infant control, response decrement, and recovery as an index of visual discrimination in young infants. In F.D. Horowitz (Ed.), Visual Attention, Auditory Stimulation, and Language Discrimination in Young Infants. Monographs of the Society for Research in Child Development, 39, (5-6, Serial No. 158), pp. 52-62.

Lavoie, C., & Desrochers, S. (2002). Visual habituation at five months: Short-term reliability of measures obtained with a new polynomial regression criterion. Journal of Genetic Psychology, 163(3), 261-271.

Lewis, M. (1969). A developmental study of information processing within the first three years of life: Response decrement to a redundant signal. Monographs of the Society for Research in Child Development, 34 (9), No. 133.

McCall, R.B., Hogarty, P.S., Hamilton, J.S., & Vincent, J.S. (1973). Habituation rate and the infant's response to visual discrepancies. Child Development, 44, 280-287.

Millar, W.S., & Weir, C. (1997). Relations between contingency learning and habituation in 6- to 12-month old infants: further evidence. Cahiers de Psychologie Cognitive, 16(5), 537-554.

Miller, D.J., Ryan, E.B., Short, E.J., Ries, P.G., McGuire, M.D., & Culler, M.P. (1977). Relationships between early habituation and later cognitive performance in infancy. Child Development, 48, 658-661.

Pelli, D.G. (1997). The video toolbox software for visual psychophysics: Transforming numbers into movies. Spatial Vision, 10, 437-442.

Petrinovich, L., & Widaman, K.F. (1984). An evaluation of statistical strategies to analyze repeated measures data. In H. V. S. Peeke and L. Petrinovich (Eds.), Habituation, Sensitization and Behavior. Chapter 5, pp. 155-201. Orlando, FL: Academic Press.

Quinn, P.C., Bhatt, R.S., Brush, D., Grimes, A., & Sharpnack, H. (2002). Development of form similarity as a gestalt grouping principle in infancy. Psychological Science, 13(4), 320-328.

Ramsey, F.L., & Schafer, D.W. (1997). The Statistical Sleuth: A Course in Methods of Data Analysis (pp. 343-345). Belmont, CA: Duxbury Press.

Seber, G.A.F., & Wild, C.J. (1989). <u>Nonlinear Regression</u>. New York, NY: Wiley Publishing.

Sirois, S., & Mareschal, D. (2002). Models of habituation in infancy. <u>Trends in Cognitive Sciences, 6(7),</u> 293-298.

Sokolov, E.N. (1963). <u>Perception and the Conditioned Reflex</u>. Oxford, UK: Pergamon Press.

Sokolov, E.N., Spinks, J.A., Näätänen, R., & Lyytinen, H. (2002). <u>The Orienting Response in Information Processing</u>. Mahwah, NJ: Lawrence Erlbaum Associates.

Thomas, H. (1993). Individual differences in children, studies, and statistics: Applications of empirical Bayes methodology. In M. Howe and R. Pasnak (Eds.), <u>Emerging Themes in Cognitive Development. Foundations</u>. Vol. I. Chapter 5, pp. 120-153. New York: Springer-Verlag.

Thomas, H., & Gilmore, R.O. (in press). Habituation assessment in infancy. <u>Psychological Methods</u>.Thompson, R.F., & Spencer, W.A. (1966). Habituation: A model phenomenon for the study of neuronal substrates of behavior. <u>Psychological Review, 73(1)</u>, 16-43.

Appendix A

Glossary of Notation and Symbols Used

Notation	Interpretation or Context
$\hat{}$	"hat" over symbol, denoting an estimated value from observed data
50%DC	the fifty percent decrement criterion
α_i	floor value for infant i, included in Models 1, 3 and 4
β_i	range of habituation function for infant i; in Models 1, 3 and 4
B	standardized B weight for linear regression transformation
d_i	matched difference score for infant i on post-habituation trials
δ_i	slope of habituation function for infant i; in Models 2, 3 and 4
δ_i^*	transformation of $\hat{\delta}_i$ that allows only a maximum value of 2
$E_i(t)$	random error term for infant i; mean is zero and independent of t
ε_i	proportion of variance accounted for by infant i's best-fit model
$h_i(t)$	unknown "true" habituation function for infant i
$h_i^*(t)$	infant i's best-fit habituation model
$h_{i,ph}^*(t)$	theoretical post-habituation model; functionally related to $h_i^*(t)$
$h_{1i}(t)$	one parameter habituation model (i.e., Model 1)
$h_{2i}(t)$	two parameter habituation model (i.e., Model 2)
$h_{3i}(t)$	three parameter habituation model (i.e., Model 3)
$h_{4i}(t)$	four parameter habituation model (i.e., Model 4)
k	number of parameters in model, or model number
μ_i	trial of maximum looking time for infant i, included in Model 4
n	number of infants in study
p	proportion or probability
$\hat{\rho}$	estimated correlation between two random variables (Pearson's r)
SEM	standard error of the mean

$\hat{\sigma}_{\hat{\theta}_i}$	bootstrap-estimated standard error of $\hat{\theta}$ for infant i
$\hat{\sigma}^2_{ik}$	error variance for infant i under model k; used to estimate BIC
$\hat{\sigma}^2_{h_i^*}$	error variance under best-fit habituation model for infant i
t	trial
t_c	current trial
T_i	total number of habituation trials for infant i
θ	generic model parameter (e.g., α_i, β_i, or δ_i as the case may be)
v	number of independent infant visits
w_i	maximum estimated looking time for infant i ($w_i = \alpha_i + \beta_i$)
$\overline{y}_{bi}$	baseline average for infant i
$\overline{y}_{ci}$	current rolling average for infant i
$y_i(t)$	observed looking time for infant i on trial t
z_i	dishabituation score for infant i
$z_{\hat{\theta}_i}$	stability z score for infant i on $\hat{\theta}_i$

Appendix B

MATLAB Code for Estimating 3-Parameter Habituation Model

```
function f3p = fit3pb(lookVect)
%
%
%     Direct search method for identifying appropriate starting
%     values for parameter estimation algorithm
%


t=[1:length(lookVect)];%defines vector of trials, t, of same length as
%                        input vector
lowest=1e13;%arbitrarily high starting point

u=zeros(20,20,35);%declares a 3-d matrix of zeros for residuals

alpha=[0:(min(lookVect)+6)/19:min(lookVect)+6];%alpha search interval

beta=[0:(max(lookVect)+20)/19:max(lookVect)+20];%beta search interval

delta=[0:10/34:10];% delta search interval

for i=1:20
   for j=1:20
         for k=1:35
                q=alpha(i);%nested loops that evaluate parameter combos
                r=beta(j);
                s=delta(k);
                 z= r^2 * exp( -(s)^2.*(t-1).^2) + q^2;
     %evaluate the model using this set of parameters at times, t
     resid=sum((z-lookVect).^2);%calculate summed squared residuals
%                     residuals between model and observed data
             u(i,j,k)=resid;%assigns value of residuals
%                      for this loop to the residual matrix
              if resid<lowest%compares residual to previous low
%                                         value and takes the lower
                    lowest=resid;
                    x0(1) = q;
%gives parameter estimates associated with this lowest residual triplet
                    x0(2) = r;
                    x0(3) = s;
                  end % if resid
              end % for k
          end % for j
       end % for i

       x0; % vector of starting points for curvefit estimates

       x = curvefit('myfun', x0, t, lookVect);
       % estimates best parameters for 'myfun'
       %(3 parameter habituation function), given starting values x0
       % t is a trial vector from 1 to length(lookVect)
```

```
x = x.^2;

expected = x(2) * exp ( - (x(3)).*((t - 1).^2)) + x(1);
% calculating resids and var.
newresids = (lookVect - expected);
ervar = sum (newresids.^2)/(length(lookVect) - 3);
Hab3BIC = BIC(length(lookVect), 3, ervar);
% calculates BIC for 3-p model

f3p(1) = Hab3BIC;
f3p(2) = x(1);
f3p(3) = x(2);
f3p(4) = x(3);
```

Appendix C

MATLAB Code for Evaluating Stability of Fits with Additional Observations

```
function fs2 = fitsequence2(lookVect)

% fitsequence:  given a dataset of looking times and a
% selection of  model type,finds the best parameter estimates for
% the data set, each time using one more observation from the
% dataset.  examines stability of parameter estimates
% each time additional observation is collected.
%  outputs a matrix with rows equal to number of observations.
%  syntax = fitsequence(lookVect, modelType)
%  lookVect:   n x 1 vector - containing observed looking times
%  type 1: 1-parameter non-Habituating Model
%  type 2: 2-parameter Thomas & Gilmore Model (floor = 0)
%  type 3: 3-parameter Thomas & Gilmore Model
%  type 4: 4-parameter Thomas & Gilmore non-Monotonic Model
%  column 1 = BIC associated with model (requires p + 1
%  observations, where p = # estimated parameters)
%  columns 2- end show the parameter estimates at that point

% Following code checks to see if the data is a vector
lookVectSize = size(lookVect);
            if (lookVectSize(1) > 1)
                if (lookVectSize(2) > 1)

                    error('Looking time data must be vector .');
                else
                    lookVect = lookVect';
                end % if

            end % if

%       Initialize variables

t = [1:length(lookVect)];
fs2 = zeros(length(lookVect), 4);
fid = fopen('Sequences.xls', 'w');


for sequence = 1:length(lookVect);
      vector = lookVect(1:sequence);
      fa(sequence, :) = fit1p(vector);
      fs2(sequence, 1) = fa(sequence, 1);

end


for sequence = 1:length(lookVect);
      vector = lookVect(1:sequence);
      fb(sequence, :) = fit2pb(vector);
      fs2(sequence, 2) = fb(sequence, 1);
```

```
end

for sequence = 1:length(lookVect);
     vector = lookVect(1:sequence);
     fc(sequence, :) = fit3pb(vector);
     fs2(sequence, 3) = fc(sequence, 1);
end

for sequence = 1:length(lookVect);
     vector = lookVect(1:sequence);
     fd(sequence, :) = fit4pb(vector);
     fs2(sequence, 4) = fd(sequence, 1);

end

fid = fopen('Sequences.xls', 'w');

fprintf(fid, '%8.4f %8.4f %8.4f %8.4f %8.4f %8.4f %8.4f %8.4f
%8.4f %8.4f %8.4f %8.4f %8.4f %8.4f %8.4f %8.4f\n', lookVect');
fprintf(fid, '\n');
fprintf(fid, '%8.4f %8.4f %8.4f %8.4f\n', fs2');
fclose(fid);


end          %switch
```

Appendix D

Matlab Code for Bootstrapped Oneway ANOVA

```
function boot3 = bootoneway3(group1, group2, group3, bootreps)

% bootoneway3:  with 3 vectors each containing independent samples,
%               uses bootstrap to simulate the p-level associated with
%               the observed F-statistic for a oneway analysis of
%               variance under the conditions of the null hypothesis.
%               In other words, uses the bootstrap to calculate
%               significance levels, thus avoiding potential pitfalls
%               due to failure to meet distributional assumptions.
%               Does not assume equal variances or sample sizes.
%
%               syntax = bootoneway3(group1, group2, group 3, bootreps)
%               group1: n x 1 vector - sample of size n from group 1
%               group2: m x 1 vector - sample of size m from group 2
%               group3: o x 1 vector - sample of size o from group 3
%               bootreps:  # of bootstrap reps. to conduct (1000 min.)
%               outputs vector containing the three observed means, the
%               observed F-value, and simulated p-value (ASL)
%               MATLAB code written August 17, 2003 by mpd

% checks to see if the Group 1 data to be tested is a vector
group1Size = size(group1);
            if (group1Size(1) > 1)
                if (group1Size(2) > 1)

                    error('Group 1 data must be 1 x n vector');
                else
                    group1 = group1';
                end % if

            end % if

% checks to see if the Group 2 data to be tested is a vector
group2Size = size(group2);
            if (group2Size(1) > 1)
                if (group2Size(2) > 1)

                    error('Group 2 data must be a 1 x n vector');
                else
                    group2 = group2';
                end % if

            end % if

% checks to see if the Group 3 data is a vector
group3Size = size(group3);
            if (group3Size(1) > 1)
                if (group3Size(2) > 1)

                    error('Group 3 data must be a 1 x n vector');
```

```
                else
                    group3 = group3';
                end % if

            end % if

%       Initialize variables
lg1= length(group1); % number of elements in vector group 1
lg2= length(group2); % number of elements in vector group 2
lg3= length(group3); % number of elements in vector group 3

avgg1 = mean(group1); % observed group sample means
avgg2 = mean(group2);
avgg3 = mean(group3);

overallmean = (sum(group1)+sum(group2)+sum(group3))/(lg1+lg2+lg3);
% computes the grand mean

SSTOT = sum((group1-overallmean).^2) + sum((group2-overallmean).^2) +
sum((group3-overallmean).^2);%total sum of squares
DFTOT = lg1+lg2+lg3-1;%total degrees of freedom
SSERROR = sum((group1-avgg1).^2) + sum((group2-avgg2).^2) + sum((group3
- avgg3).^2);%error term sum of squares
DFERROR = lg1+lg2+lg3-3;% degrees freedom error term
SSTREAT = SSTOT - SSERROR;
DFTREAT = 2; % for a oneway with three groups this is constant at 2

OBSF = (SSTREAT/DFTREAT)/(SSERROR/DFERROR);

adjgroup1 = group1 - avgg1 + overallmean;
adjgroup2 = group2 - avgg2 + overallmean;
adjgroup3 = group3 - avgg3 + overallmean;
%adjusts observed samples to the conditions of the null hypothesis-
% i.e., that the three group means are equal to the grand mean


g1bootmatrix = zeros(bootreps, lg1);
g2bootmatrix = zeros(bootreps, lg2);
g3bootmatrix = zeros(bootreps, lg3);
bootfmatrix = zeros(bootreps, 8);
% declares matrices in which to store boot samples and boot statistics

for bsrep = 1:bootreps
% each row of bootsample matrices are a boot sample with replacement
% from the observed sample for each group

    g1bootmatrix(bsrep,:) = swr(adjgroup1, lg1);
        g1boot = g1bootmatrix(bsrep,:);
    g2bootmatrix(bsrep,:) = swr(adjgroup2, lg2);
        g2boot = g2bootmatrix(bsrep,:);
    g3bootmatrix(bsrep,:) = swr(adjgroup3, lg3);
        g3boot = g3bootmatrix(bsrep,:);
    bootrepgrandmean= (sum(g1boot)+sum(g2boot) + sum(g3boot))/(DFTOT+1);

    bootfmatrix(bsrep, 1) = mean(g1bootmatrix(bsrep,:)); %avg boot grp1
        g1bar = bootfmatrix(bsrep, 1);
    bootfmatrix(bsrep, 2) = mean(g2bootmatrix(bsrep,:)); %avg boot grp2
        g2bar = bootfmatrix(bsrep, 2);
    bootfmatrix(bsrep, 3) = mean(g3bootmatrix(bsrep,:)); %avg boot grp3
        g3bar = bootfmatrix(bsrep, 3);
    bootfmatrix(bsrep, 4) =  sum((g1boot - bootrepgrandmean).^2) +
        sum((g2boot - bootrepgrandmean).^2) + sum((g3boot -
```

```
      bootrepgrandmean).^2); % SS Total for this boot sample
      bootsstot = bootfmatrix(bsrep, 4);
   bootfmatrix(bsrep, 5) = sum((g1boot-g1bar).^2) + sum((g2boot-
      g2bar).^2) + sum((g3boot-g3bar).^2);%SS error for boot sample
      bootsserror = bootfmatrix(bsrep, 5);
   bootfmatrix(bsrep, 6) = bootsstot - bootsserror; %SS treatment
      bootsstreat = bootfmatrix(bsrep, 6);
   bootfmatrix(bsrep, 7) = (bootsstreat/DFTREAT)/(bootsserror/DFERROR);
   % F-statistic for this boot sample
   bootf = bootfmatrix(bsrep, 7);

      if bootfmatrix(bsrep, 7) >= OBSF;
           bootfmatrix(bsrep, 8) = 1;
% counter to see what proportion of the total number of boot F-ratios
% are at least as big as the observed value from original sample
      end %if

end % for loop

boot3(1) =avgg1;
boot3(2) = avgg2;
boot3(3) = avgg3;
boot3(4) = OBSF;
boot3(5) = sum(bootfmatrix(:,8))/bootreps;
```

Matlab Code for Bootstrapped Follow-up Group Mean Comparisons

```
function btt3 = boottukey3(group1, group2, group3)


% checks to see if the Group 1 data is a vector
group1Size = size(group1);
            if (group1Size(1) > 1)
                if (group1Size(2) > 1)

                    error('Group 1 data must be a vector.');
                else
                    group1 = group1';
                end % if

            end % if


% checks to see if the Group 2 data is a vector
group2Size = size(group2);
            if (group2Size(1) > 1)
                if (group2Size(2) > 1)

                    error('Group 2 data must be a vector.');
                else
                    group2 = group2';
                end % if

            end % if

% checks to see if the Group 3 data is a vector
group3Size = size(group3);
            if (group3Size(1) > 1)
                if (group3Size(2) > 1)

                    error('Group 3 data must be a vector.');
                else
                    group3 = group3';
                end % if

            end % if

%       Initialize variables
lg1= length(group1); % number of elements in vector group 1
lg2= length(group2); % number of elements in vector group 2
lg3= length(group3); % number of elements in vector group 3

avgg1 = mean(group1); % observed group sample means
avgg2 = mean(group2);
avgg3 = mean(group3);
```

```
g1g2dif = avgg1-avgg2;% mean group differences
g1g3dif = avgg1-avgg3;
g2g3dif = avgg2-avgg3;

bootreps = 3600;% big enough number easily divisible by 6

g1bootmatrix = zeros(bootreps, lg1);
% declares matrices in which to store boot samples and statistics
g2bootmatrix = zeros(bootreps, lg2);
g3bootmatrix = zeros(bootreps, lg3);
boottukmatrix = zeros(bootreps, 3);

for bsrep = 1:bootreps
      g1bootmatrix(bsrep,:) = swr(group1, lg1);
            % each row of the bootsample matrices are a boot
            % sample with replacement from the observed groups
        bootg1avg = mean(g1bootmatrix(bsrep,:));
      g2bootmatrix(bsrep,:) = swr(group2, lg2);
          bootg2avg = mean(g2bootmatrix(bsrep,:));
      g3bootmatrix(bsrep,:) = swr(group3, lg3);
          bootg3avg = mean(g3bootmatrix(bsrep,:));

      boot12dif = bootg1avg - bootg2avg;
      boot13dif = bootg1avg - bootg3avg;
      boot23dif = bootg2avg - bootg3avg;
      %group mean differences for each boot sample

      boottukmatrix(bsrep, 1) = boot12dif;
      boottukmatrix(bsrep, 2) = boot13dif;
      boottukmatrix(bsrep, 3) = boot23dif;

end % for loop

sortboot = sort(boottukmatrix);

btt3(1,1) =avgg1;
btt3(1,2) =avgg2;
btt3(1,3) =avgg3;
btt3(2,1) = g1g2dif ;
btt3(2,2) = g1g3dif;
btt3(2,3) =  g2g3dif;
btt3(3,1) =  sortboot(31,1);
btt3(3,2) =  sortboot(31,2);
btt3(3,3) =  sortboot(31,3);
btt3(4,1) =  sortboot(3570,1);
btt3(4,2) =  sortboot(3570,2);
btt3(4,3) =  sortboot(3570,3);
```

VITA

Michael P. Dahlin   May 2004

Contact Information

205 Hillcrest Drive

Selah, WA 98942

(509) 697-4721

mdahlin@psu.edu

Education

Graduate Studies

Doctor of Philosophy

The Pennsylvania State University; University Park, PA 16802

Degree Awarded: May 2004

Major: Developmental Psychology

Minor: Statistics

Thesis Title: *Infant Visual Habituation Modeled by Non-linear Regression*

Chair: Hoben Thomas, Ph.D.

Master of Science

Western Washington University; Bellingham, WA 98225

Degree Awarded: August 1997

Major: General Psychology

Thesis Title: *Parental discourse styles: Stability and relationship to child language skills.*

Chair: Catherine Crain-Thoreson, Ph.D.

Undergraduate Studies

Bachelor of Arts

Pomona College; Claremont, CA 91711

Degree Awarded: May 1990